Two algorithms for the discrete time approximation of Markovian backward stochastic differential equations under local conditions

Plamen Turkedjiev

Two algorithms for the discrete time approximation of Markovian backward stochastic differential equations under local conditions

Plamen Turkedjiev*
Centre de Mathématiques Appliquées
Ecole Polytechnique and CNRS
Route de Saclay
F 91128 Palaiseau cedex, France

20th August, 2014

ABSTRACT: Two discretizations of a class of locally Lipschitz Markovian backward stochastic differential equations (BSDEs) are studied. The first is the classical Euler scheme which approximates a projection of the processes Z, and the second a novel scheme based on Malliavin weights which approximates the marginals of the process Z directly. Extending the representation theorem of Ma and Zhang [MZ02] leads to advanced a priori estimates and stability results for this class of BSDEs. These estimates are then used to obtain competitive convergence rates for both schemes with respect to the number of points in the time-grid. The class of BSDEs considered includes Lipschitz BSDEs with fractionally smooth terminal condition, thus extending the results of [GM10], quadratic BSDEs with bounded, Hölder continuous terminal condition (for bounded, differentiable volatility), and BSDEs related to proxy methods in numerical analysis.

KEYWORDS: Backward stochastic differential equation, approximation schemes, Malliavin calculus, representation theorem, a priori estimates.

MSC 2010: 60H35, 65C30, 60H07, 60H10.

1 Introduction

▶ **Framework.** Backward stochastic differential equations play an important role in the theory of mathematical finance, stochastic optimal control, and partial differential equations. In this paper, we study two discrete-time approximations of the for the so-called *locally Lipchitz* Markovian backward stochastic differential equation (BSDE). The purpose is to determine the error induced by these approximations under suitable norms. The first is the well-established Euler scheme for BSDEs, and the second is a novel scheme we call the Malliavin weights scheme for BSDEs. Let $T >$

*This research benefitted from the support of Chair Financial Risks of the Risk Foundation, the FiME Laboratory, and the Chair Finance and Sustainable Developement, under the aegis of Louis Bachelier Finance and Sustainable Growth laboratory, a joint initiative with Ecole Polytechnique. A significant part of the author's research was done at Humboldt University. Email: turkedjiev@cmap.polytechnique.fr. The author would like to thank in particular Dirk Becherer and Emmanuel Gobet for their feedback and spotting some early errors in the manuscript.

0 be a fixed terminal time and $(\Omega, \mathcal{F}_T, \{\mathcal{F}_t\}, \mathbb{P})$ a filtered probability space, where $\{\mathcal{F}_t : 0 \le t \le T\}$ is the filtration generated by a q-dimensional $(q \ge 1)$ Brownian motion W and satisfying the usual conditions of right-continuity and completeness. We look to approximate the $\mathbb{R} \times (\mathbb{R}^q)^\top$-valued, predictable process (Y, Z) solving the BSDE

$$Y_t = \Phi(X_T) + \int_t^T f(s, X_s, Y_s, Z_s)ds - \int_t^T Z_s dW_s. \tag{1.1}$$

Here, $(\mathbb{R}^q)^\top$ is the space of q-dimensional, real valued row vectors; X is an $\mathbb{R}^d$-valued $(1 \le d \le q)$ diffusion; and $\Phi : \mathbb{R}^d \to \mathbb{R}$ and $f : [0,T) \times \mathbb{R}^d \times \mathbb{R} \times (\mathbb{R}^q)^\top \to \mathbb{R}$ are deterministic functions that are termed the *terminal condition* and *driver*, respectively. We focus on the setting in which the terminal condition Φ is in the space of **fractionally smooth** functions $\mathbf{L}_{2,\alpha}$ for parameter $\alpha \in (0,1]$ - see $(\mathbf{A}_\Phi)$ in Section 1.2 for details - and the driver is **locally Lipschitz continuous in** (x, y, z) and **locally bounded at** 0 in the sense that there exist exponents $\theta_L, \ \theta_X, \ \theta_c \in (0,1]$, finite constants $L_f, L_X, C_f \ge 0$, such that, for all $t \in [0,T)$ and $(x,y,z), (x',y',z') \in \mathbb{R}^d \times \mathbb{R} \times (\mathbb{R}^q)^\top$,

$$\begin{aligned} |f(t,x,y,z) - f(t,x',y',z')| &\le L_f \frac{|y-y'| + |z-z'|}{(T-t)^{(1-\theta_L)/2}} + L_X \frac{|x-x'|}{(T-t)^{1-\theta_X/2}}, \\ |f(t,x,0,0)| &\le \frac{C_f}{(T-t)^{1-\theta_c}}. \end{aligned} \tag{1.2}$$

Furthermore, X solves a time-inhomogeneous stochastic differential equation (SDE) with suitable coefficients; see $(\mathbf{A}_{\mathbf{b},\sigma})$ in Section 1.2. The existence and uniqueness of this class of BSDEs – given in Section 2.3 – follows from [FJ12, Theorem 3.2]. Below, we show that this class of BSDEs includes a section of the important quadratic BSDEs, and also BSDEs related to so-called proxy schemes used for numerical methods, so it is of interest to find good discrete-time approximations for such BSDEs. We note that fully implementable algorithms – admitting the full generality of the assumptions considered in this paper – based on the Euler and Malliavin weights schemes have been studied in detail in [GT13b][GT13a] respectively, but, to the best of our knowledge, this is the first paper considering the discretization error under the full generality of the local conditions.

► **Summary of results.** In the spirit of [GM10], we make use of non-uniform time-grids $\{\pi_N^{(\beta)} := \{0 = t_0^{(N)} < \ldots < t_N^{(N)} = T\} \ : \ N \ge 1\}$ whose parameter $\beta \in (0,1]$ determines the time-points $t_i^{(N)} := T - T(1 - i/N)^{1/\beta}$. As in [GM10], the use of these time-grids appears to substantially reduce the error due to disctretization.

The first approximation, studied in Section 3, is the so-called Euler scheme for BSDEs:

$$\begin{aligned} Y_N^{(N)} &:= \Phi(X_T), \quad Z_i^{(N)} := \frac{1}{t_{i+1}^{(N)} - t_i^{(N)}} \mathbb{E}[Y_{i+1}^{(N)} (W_{t_{i+1}^{(N)}} - W_{t_i^{(N)}})^\top | \mathcal{F}_{t_i^{(N)}}], \\ Y_i^{(N)} &:= \mathbb{E}[Y_{i+1}^{(N)} + f(t_i^{(N)}, X_{t_i^{(N)}}, Y_{i+1}^{(N)}, Z_i^{(N)})(t_{i+1}^{(N)} - t_i^{(N)}) | \mathcal{F}_{t_i}] \end{aligned} \tag{1.3}$$

for each $i \in \{0, \ldots, N-1\}$. The random variable $Z_i^{(N)}$ is a discretization of the projection $(t_{i+1}^{(N)} - t_i^{(N)})\tilde{Z}_{t_i} := \mathbb{E}[\int_{t_i^{(N)}}^{t_{i+1}^{(N)}} Z_s ds | \mathcal{F}_{t_i^{(N)}}]$. This approximation has been frequently studied: [Zha04][BT04][GL07] among others, in the setting where the terminal condition Φ and the driver are uniformly Lipschitz continuous (i.e. $\theta_L = 1$); [GM10] in the setting of the fractionally smooth Φ but uniformly Lipschitz continuous driver; [IDR10][Ric11] in the setting of bounded Lipschitz (resp. Hölder) continuous Φ and quadratic driver; and [Ric12] in the setting of possibly unbounded (locally) Lipschitz continuous Φ and (super-)quadratic driver. Typically, the discretization error of the Euler scheme is measured by

$$\mathcal{E}(N) := \max_{0 \le i < N} \mathbb{E}[|Y_{t_i^{(N)}} - Y_i^{(N)}|^2] + \sum_{i=0}^{N-1} \int_{t_i^{(N)}}^{t_{i+1}^{(N)}} \mathbb{E}[|Z_t - Z_i^{(N)}|^2] dt. \tag{1.4}$$

We show in Theorem 3.3 that if $\beta < (2\gamma) \wedge \alpha$, where $\gamma := (\frac{\alpha}{2} \wedge \theta_c + \frac{\theta_L}{2}) \wedge \theta_c$, then

$$\mathcal{E}(N) \leq CN^{-1}\mathbf{1}_{[1,2]}(\alpha + \theta_L) + CN^{-2\gamma}\mathbf{1}_{(0,1)}(\alpha + \theta_L).$$

The optimal error bound $O(N^{-1})$ is obtained if $\alpha+\theta_L \geq 1$. This rate is optimal in the sense that it is the same as the rate of convergence obtained in [GM10, Theorem 3.2] in the uniformly Lipschitz driver setting ($\theta_L = 1$). This result can be complimented under the additional assumption that the terminal condition Φ is θ_Φ-Hölder continuous: in Theorem 4.5, we show that if $\beta < (2\gamma) \wedge \alpha \wedge \theta_L$, then

$$\mathcal{E}(N) \leq CN^{-1}\mathbf{1}_{[1,4]}(\theta_\Phi + \beta + 2\gamma) + CN^{-2\gamma}\mathbf{1}_{(0,1)}(\theta_\Phi + \beta + 2\gamma).$$

Now $\theta_\Phi + \beta + 2\gamma \geq 1$ is sufficient to obtain the optimal convergence rate $O(N^{-1})$. Although the complex relationship between θ_Φ, α and γ make it difficult to compare the two results in full generality, the latter result relaxes the constraint $\alpha + \theta_L \geq 1$ in order to obtain the optimal error bound $O(N^{-1})$ if $\theta_c \geq 1/2$ – see (1.2) to recall the definition of θ_c.

The second approximation, studied in Section 5, is the so-called Malliavin weights scheme. Rather than approximating the projections of the process Z, this algorithm is used to approximate the version of Z, determined by the Malliavin integration-by-parts formula of Theorem 2.16, at the points of the time grid directly: for each $N \geq 1$, set

$$\bar{Y}_N^{(N)} := \Phi(X_T), \quad \bar{Y}_i^{(N)} := \mathbb{E}[\Phi(X_T) + \sum_{j=i}^{N-1} f(t_j^{(N)}, X_{t_j^{(N)}}, \bar{Y}_{j+1}^{(N)}, \bar{Z}_j^{(N)})(t_{j+1}^{(N)} - t_j^{(N)})|\mathcal{F}_{t_i^{(N)}}],$$
$$\bar{Z}_i^{(N)} := \mathbb{E}[\Phi(X_T)H_N^i + \sum_{j=i+1}^{N-1} f(t_j^{(N)}, X_{t_j^{(N)}}, \bar{Y}_{j+1}^{(N)}, \bar{Z}_j^{(N)})H_j^i(t_{j+1}^{(N)} - t_j^{(N)})|\mathcal{F}_{t_i^{(N)}}] \tag{1.5}$$

for $i \in \{0, \ldots, N-1\}$, where $(H_j^i)_{i,j}$ is a suitable random variable. Due to the connection between BSDEs and quasilinear partial differential equations (PDEs) – see [Ric12][CD12] and references therein – it may be of interest to approximate the marginals of the process Z rather than the projections. Other schemes that make use of Malliavin calculus are available [BL13][HNS11], but this is, to the best of our knowledge, the first scheme which makes use of the Malliavin integration-by-parts formula (Theorem 2.16). Convergence results are given – for weaker norms than those used in $\mathcal{E}(N)$ for the Euler scheme – in Theorem 5.5. Although one is able to prove results under stronger norms than for the Euler scheme, there are several disadvantages (regardless of the norm used to measure the error) of the Malliavin weights scheme over the Euler scheme. Our results are proven under stronger conditions than for the Euler scheme because the use of stronger a priori estimates – Proposition 4.2 – is essential in the proof: one requires that either the terminal condition has exponential moments or that it is Hölder continuous. We have not yet been able to weaken the conditions on these a priori estimates. One also requires a greater constraint $\beta \leq \gamma \wedge \theta_L \wedge \alpha$ (where $\gamma := (\frac{\alpha}{2} \wedge \theta_c + \frac{\theta_L}{2}) \wedge \theta_c$) on the time-grid than for the Euler scheme. The rate of convergence again depends on the parameters $(\alpha, \theta_L, \theta_c, \beta)$. In the more general setting of exponential moments on the terminal condition, $\beta + 2\gamma \geq 1$ is required for the optimal error bounds $O(N^{-1})$, whereas in the setting of θ_Φ- Hölder continuous terminal condition, $\beta + \theta_\Phi + 2\gamma \geq 1$ is sufficient. One may ask, given the additional constraints, why it is of interest to study the Malliavin weights scheme over the Euler scheme? The reason has to do with the approximation of the conditional expectation. It is shown in [GT13a] that, using Monte Carlo least-squares regression to approximate the conditional expectation, one can theoretically gain an order one improvement with respect to the number of time-steps N on the algorithm complexity using the Malliavin weights scheme compared to the multi-step forward implementation of the Euler scheme [GT13b]. Such a complexity reduction is substantial, given that N may be very large.

In order to obtain the results on discretization, we extend some basic tools from the literature of BSDEs. These results are interesting in their own right. Firstly, we extend stability estimates for Lipschitz BSDEs to the class of BSDEs satisfying local Lipschitz continuity and boundedness conditions (1.2). This enables us to make estimates on the basis of constructing approximating sequences, a key technique used throughout the paper. A natural consequence of stability estimates are a priori estimates, which we also frequently require. These results are contained in Section 2.4. Secondly, we obtain dynamical representations of the process Z_t in the form of the product $U_t\sigma(t, X_t)$, where (U, V) is the solution of a linear BSDE. Such representations are very valuable for making estimates on the increments $\mathbb{E}[|Z_t - Z_s|^2]$, because one can make use of a priori estimates on the linear BSDE and the process X. In fact, it is not possible to obtain the results for Z directly, but for a suitable sequence $\{Z_t^{(\varepsilon)} \ : \ \varepsilon > 0\}$ of approximating BSDEs. A priori estimates for the approximation are computed and play an important role in the overall convergence rate of the numerical schemes. To obtain this result, we extend the method and results of [GM10, Section 2], who consider the setting (1.2) with $\theta_L = \theta_c = 1$ only, to our more general setting. The key results are contained in Lemma 2.9. Thirdly, we extend the classical representation theorem of Ma and Zhang [MZ02, Theorem 4.2] for the Z process to our class of BSDEs. This theorem is proved in Section 2.5 and is a key result in this paper. One the one hand, it is the basis for the Malliavin weights scheme. On the other hand, we use the representation theorem to obtain stability estimates directly on the marginals of the process Z – see Proposition 2.12 – which are key to the analysis. These stability estimates lead in turn to a priori estimates of the form

$$|Z_t| \leq C\sqrt{(T-t)^{-1}\mathbb{E}_t[|\Phi(X_T) - \mathbb{E}_t[\Phi(X_T)]|^2]} + C(T-t)^{\theta_c - 1/2} + C\mathbb{E}_t[\Phi(X_T)^2]^{1/2}(T-t)^{\theta_L/2}$$

for all $t \in [0, T)$ almost surely. Such estimates are, to the best of our knowledge, novel and allow us to study the impact of the regularity of the terminal condition on a priori estimates – see Proposition 2.13. Finally, in Proposition 4.2, we obtain a priori estimates for the process $V_t^{(\varepsilon)}$ – the solution $(U^{(\varepsilon)}, V^{(\varepsilon)})$ to the linear BSDE such that the approximating BSDE solution satisfies $Z_t^{(\varepsilon)} = U_t^{(\varepsilon)}\sigma(t, X_t)$ – under additional regularity conditions on the terminal condition. These estimates are essential to analyse the error due to the Malliavin weight scheme. Rather than considering a second Malliavin derivative of the process Y_t, as for example do [CD12], we make use of a functional representation that comes from the Markov property of X and determine regularity properties of the said functional representation. A consequence of this is the Lipschitz continuity of the functional representation of the process Z_t under suitable conditions – see Corollary 4.3. To our knowledge, this result is novel. Since regularity properties are very useful for the calibration of numerical schemes – see for example [GT13b, Section 4.4] – this result may have some impact on reducing the cost of fully implementable algorithms.

► **Contributions to quadratic BSDEs and proxy methods.** We consider the setting where Φ is a bounded, θ_Φ-Hölder continuous function. To make the contributions of the numerical results in this paper clearer, we consider two important examples. Note that these examples have also been given some attention in [GT13b, Section 2]. We emphasize that the forward process X is a diffusion with bounded, twice continuously differentiable coefficients, whose partial derivatives are bounded and Hölder continuous; this assumption stands throughout this paper – see $(\mathbf{A_{b,\sigma}})$.

Quadratic BSDEs have powerful applications in financial mathematics, for example to solve utility optimization problems in incomplete markets [REK00][HIM05]. Let $q = d$ and the measurable function $F : [0, T) \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ satisfy

$$|F(t, x, y, z)| \leq c\ (1 + |y| + |z|^2),$$
$$|F(t, x, y, z) - F(t, x, y', z')| \leq c\ (1 + |z| + |z'|)(|y - y'| + |z - z'|).$$

It is known [DG06] that the solution (Y, Z) of the BSDE with terminal condition Φ and driver $F(t,x,y,z)$ exists and is unique and that there is a constant $\theta \in (0,1]$ and finite $C_u > 0$ such that $|Z_t| \leq C_u(T-t)^{(\theta-1)/2}$ for all $t \in [0,T)$ almost surely. This implies that (Y,Z) also solves the BSDE under local conditions with terminal condition Φ and driver $f(t,x,y,z) := F(t,x,y,\mathcal{T}_{C_u(T-t)^{(\theta-1)/2}}(z))$, where $\mathcal{T}_L(z) := (-L \vee z_1 \wedge L, \ldots, -L \vee z_q \wedge L)$. Indeed, $C_f = c$, $\theta_c = 1$, $L_f = c(T^{(1-\theta)/2} + 2\sqrt{d}C_u)$, and $\theta_L = \theta$. The terminal condition is fractionally smooth with parameter α at least as large as θ_Φ - see Remark 1.3. It is shown in Corollary 2.13 that $|Z_t| \leq C(T-t)^{(\theta_\Phi-1)/2}$, so θ_L is at least as large as θ_Φ. Therefore, the error $\mathcal{E}(N)$ of the Euler scheme is bounded above by $C_\beta N^{-1}\mathbf{1}_{[1,4]}(3\theta_\Phi+\beta) + CN^{-2\theta_\Phi}\mathbf{1}_{(0,1)}(3\theta_\Phi+\beta)$ for any $\beta < \theta_\Phi$. In [Ric11], the Euler scheme for bounded, Hölder continuous is also considered, but with a different non-uniform time-grid and a transformation of the terminal condition; there is a further modelling difference in that the author requires no uniform elliptic condition, but sacrifices state-dependence in the volatility matrix. The author obtains a rate of convergence $C_\eta N^{\eta-\theta_\Phi}$ for any $\eta > 0$, so we have obtained an improvement in this work; This improvement is likely due to the use of the time-grids $\pi_N^{(\beta)}$ in our scheme – indeed, [GM10] show a rate of convergence $O(N^{-\alpha})$ in the uniformly Lipschitz continuous driver setting if only a uniform time-grid is used. It is important to remark that this work is a complement to the recent papers [Ric12][CR14], in which the authors consider weaker assumptions on the drift and the volatility of the SDE – only Lipschitz continuity and linear growth are required – however stronger assumptions are required on the terminal function Φ, which must be locally Lipschitz continuous.

Next we consider a particular instance of the proxy method. Let $F(t,x,y,z)$ satisfy (1.2) with exponents $\theta_{L,F} \leq 1$, $\theta_{X,F} = 1$ and $\theta_{c,F} = 1$, and constants L_F, $L_{F,X}$ and C_F. Let $(\mathcal{Y}, \mathcal{Z})$ satisfy the BSDE with terminal condition Φ and driver $F(t,x,y,z)$. Let the function $\bar{F}(t,x,y,z)$ satisfies (1.2) with exponents $\theta_{L,\bar{F}} = \theta_{X,\bar{F}} = \theta_{c,\bar{F}} = 1$, and constants $L_{\bar{F}}$, $L_{\bar{F},X}$ and $C_{\bar{F}}$, and $\bar{\Phi}(x)$ is θ_Φ-Hölder continuous and suppose that the parabolic PDE

$$0 = \partial_t v + \bar{\mathcal{L}}_{t,x} v + \bar{F}(t,x,v(t,x),\nabla_x v(t,x)\sigma(t,x)), \qquad v(T,x) = \bar{\Phi}(x)$$

has a unique strong solution v, and, for every $t \in [0,T)$, the k-th order $(k \leq 3)$ partial derivatives in x of v are bounded by $C_u(T-t)^{(\theta_\Phi-k)/2}$. We assume also that the parabolic operator $\bar{\mathcal{L}}_{t,x}$ satisfies the property that, for any $i \in \{1,\ldots,d\}$, $\|\partial_{x_i}\{\bar{\mathcal{L}}_{t,x} - \mathcal{L}_{t,x}\}v(t,\cdot)\|_\infty \leq C_u(T-t)^{(\theta_\Phi-2)/2}$, where $\mathcal{L}_{t,x}$ is the parabolic operator given by

$$\mathcal{L}_{t,x}u(t,x) := \left\{ \frac{1}{2}\sum_{i,j=1}^d (\sigma(t,x)\sigma(t,x)^\top)_{i,j} \frac{\partial^2}{\partial_{x_i}\partial_{x_j}} + \sum_{i=1}^d b_i(t,x)\frac{\partial}{\partial_{x_i}} \right\} u(t,x);$$

this is stronger than the previous assumption on the third order partial derivatives of $v(t,\cdot)$, which asks for the upper bound $C_u(T-t)^{(\theta_\Phi-3)/2}$. Then $(Y_t, Z_t) := (\mathcal{Y}_t - v(t,X_t), \mathcal{Z}_t - \nabla_x v(t,X_t)\sigma(t,X_t))$ solves a BSDE with terminal condition $\Phi(x) - \bar{\Phi}(x)$ and driver

$$f(t,x,y,z) := F(t,x,v(t,x)+y,\nabla_x v(t,x)\sigma(t,x)+z) - \bar{F}(t,x,v(t,x),\nabla_x v(t,x)\sigma(t,x)) + (\mathcal{L}_{t,x} - \bar{\mathcal{L}}_{t,x})v(t,x).$$

The driver $f(t,x,y,z)$ satisfies (1.2) with exponents $\theta_L = \theta_{L,F}$, $\theta_X = \theta_L + \theta_\Phi - 1$, $\theta_c = \theta_\Phi$, and constants $L_f = L_F$, $L_X := L_F C_u + \sqrt{d}C_u T^{(1-\theta_L)/2}(1+L_{\bar{F}})$, and $C_f = \sqrt{d}(L_F + L_{\bar{F}})C_u + C_F + C_{\bar{F}}$. The idea is that it may be numerically advantageous to simulate the BSDE (Y,Z) as opposed to the original BSDE $(\mathcal{Y},\mathcal{Z})$. A simple example of a proxy is given by $\bar{\Phi}(x) \equiv \Phi(x)$, $\bar{F} \equiv 0$, and $\bar{\mathcal{L}}_{t,x}u(t,x) = \mathcal{L}_{t,x}u(t,x)$; see Lemma 2.8 for the gradient bounds. We show in Corollary 4.3 that the process (Y,Z) brought about by this proxy may lead to some regularity improvements for the process Z compared with the original process $\mathcal{Z}$. This may lead to an improvement of the numerical complexity for fully implementable algorithms that approximate the conditional expectation, where regularity is extremely important; moreover, [GT13b][GT13a] both demonstrate that there will an improvement in the constants for the error estimates when using Monte Carlo least-squares regression on this proxy compared to the same algorithm on the original BSDE $(\mathcal{Y}, \mathcal{Z})$.

► **Remarks on extensions.** In this paper, we work with one of the simplest time-inhomogeneous SDE models with stochastic volatility, which, in particular, allows us to make use of results from the theory of parabolic PDEs [Fri64] – see Lemma 2.8. The representation theorem for Z in Theorem 2.16 also makes use of the uniform ellipticity condition. Our application to quadratic BSDEs requires these conditions, and additionally that Φ is Hölder continuous and bounded, because we make use of the results of [DG06] to introduce local Lipschitz continuity. There are already several directions that may help us to avoid the uniformly elliptic condition. The results of [Kus03][CD12][Nee11], offer suitable PDE results under UFG conditions. Also, a representation theorem beyond the uniformly elliptic setting has been found by [Zha05] and [GM$^+$05] (although only for the zero driver case in the second reference). Another interesting aspect of our general results is that we require neither BMO results nor (local)-Lipschitz continuity of Φ. Combined with the connection to quadratic BSDEs already discussed here, this suggests the results of this paper may be an important stepping-stone to obtain novel representation theorems, a priori estimates, existence and uniqueness results for (super-)quadratic BSDEs with possibly unbounded and discontinuous terminal conditions. It would also be interesting to combine the results of this paper with those of [Ric12] to handle the setting of unbounded, state-dependent σ with non-Lipschitz continuous terminal condition. Unfortunately, all of these extensions are beyond the scope of this paper.

1.1 Notation and conventions

►**Time-grids.** Since each result is given for a fixed number of time-points N, we denote the points $\{t_i^{(N)}\}$ of the time-grid simply by $\{t_i\}$. Let $\Delta_i := t_{i+1} - t_i$ and $\Delta W_i := W_{t_{i+1}} - W_{t_i}$. We also suppress the superscript (N) in the Euler and Malliavin weights scheme.
► **Expectations and norms.** For $p \geq 1$, we denote by $\|\cdot\|_p$ the norms $(\mathbb{E}[|\cdot|^p])^{1/p}$; in particular, we make use of the norm $\sqrt{\mathbb{E}[|\cdot|^2]}$ denoted by $\|\cdot\|_2$.
► **Conditional expectations.** The conditional expectation $\mathbb{E}[\cdot|\mathcal{F}_t]$ is denoted by $\mathbb{E}_t[\cdot]$, and $\mathbb{E}_{t_i}[\cdot]$ is denoted $\mathbb{E}_i[\cdot]$. We make use of a conditional version of Fubini's theorem, stated in Lemma A.1. We slightly abuse notation by writing $\int_0^T \mathbb{E}_t[f_s]ds := \int_0^T F_t(\cdot, s)ds$, (likewise $\int_0^T g(\mathbb{E}_t[f_s])ds := \int_0^T g(F_t(\cdot, s))ds$ for any measurable function g) where F_t is the process defined in Lemma A.1, because we believe this notation to be somewhat clearer – in particular, this formal definition indicates more clearly that the inner integral comes from a conditional expectation than strictly mathematically correct version using the process $F_t(\cdot, s)$.
► **Lebesgue measure** For any Euclidean space E, $\mathcal{B}(E)$ denotes the Borel measurable sets in E, and the Lebesgue measure on the measurable space $\big(E, \mathcal{B}(E)\big)$ is denoted by m.
► **Processes and spaces.** For two processes X and Y in $\mathbf{L}_0([0,T] \times \Omega; \mathbb{R}^k)$, Y is said to be a version of X if $X = Y$ $m \times \mathbb{P}$-a.e. $\mathcal{P} \subset \mathcal{B}([0,T]) \otimes \mathcal{F}_T$ is the predictable σ-algebra, generated by the continuous, adapted processes, and $\mathcal{H}^2$ is the subspace of $\mathbf{L}_2([0,T] \times \Omega)$ containing only predictable processes. For $p \geq 2$, $\mathcal{S}^p$ is the subspace of $\mathcal{H}^2$ of continuous processes Y such that $\|Y\|_{\mathcal{S}^p} := (\mathbb{E}[\sup_{0 \leq s \leq T} |Y_s|^p])^{\frac{1}{p}}$ is finite for all $Y \in \mathcal{S}^p$; $\|\cdot\|_{\mathcal{S}^p}$ is a norm for this space.
► **Linear algebra** We identify the space of $k \times n$ dimensional, real valued matrices with $\mathbb{R}^{k \times n}$. $x^\top$ denotes the transpose of the vector x. I_n denotes the identity matrix in $\mathbb{R}^{n \times n}$. For any $A \in \mathbb{R}^{k \times n}$, let A_j denote the j-th column vector of A. For any vector $x \in \mathbb{R}^n$, $|x|$ is the vector 2-norm, defined

by $(\sum_{i=1}^n |x_i|^2)^{1/2}$, and for any matrix A, $|A|$ is the matrix 2-norm, defined by $\max_{|x|=1}|Ax|$, where $|Ax|$ is the vector 2-norm of the vector Ax.

► **Functions and regularity.** Let $\gamma \in (0,1]$ and $A(\cdot)$ be a function in the domain $[0,T)\times\mathbb{R}^l$ taking values in $\mathbb{R}^{k\times n}$ (resp. $\mathbb{R}^k$). We say that $A(t,\cdot)$ is γ-Hölder continuous uniformly in t with Hölder constant L_A if, for all $(x,y)\in(\mathbb{R}^l)^2$ and $t\in[0,T)$, $|A(t,x)-A(t,y)|\le L_A|x-y|^\gamma$; in the case that $\gamma=1$, we say that $A(t,\cdot)$ is Lipschitz continuous uniformly in t with Lipschitz constant L_A. Likewise, we say that $A(\cdot,x)$ is γ-Hölder continuous uniformly in x with Hölder constant L_A if, for every $(t_1,t_2)\in[0,T)^2$ and $x\in\mathbb{R}^l$, $|A(t_1,x)-A(t_2,x)|\le L_A|t_1-t_2|^\gamma$. For a given multi-index $\alpha=(i_1,\dots,i_{|\alpha|})$ with no zero entries, we define by $\partial_x^\alpha A(t,\cdot)$ the multiple derivative $\partial_{x_{i_1}}\dots\partial_{x_{i_{|\alpha|}}}A(t,\cdot)$. If $A(t,\cdot)$ takes values in $\mathbb{R}^k$ and is differentiable, we define by $\nabla_x A(t,\cdot)$ the $\mathbb{R}^{k\times l}$ valued function whose (u,v)-th component is $\partial_{x_v}A_u(t,\cdot)$. If $A(t,\cdot)$ takes values in $(\mathbb{R}^k)^\top$ and is differentiable, we define by $\nabla_x A(t,\cdot)$ the $\mathbb{R}^{l\times k}$-valued function whose (u,v)-th component is $\partial_{x_u}A_v(t,\cdot)$. Define by $\|A\|_\infty$ the infinity norm

$$\max_{u,v}\sup_{(t,x)\in[0,T)\times\mathbb{R}^l}|A_{u,v}(t,x)| \quad (\text{resp.}\ \max_u \sup_{(t,x)\in[0,T)\times\mathbb{R}^l}|A_u(t,x)|).$$

► **Mollifiers.** The following definitions will come in handy.

Definition 1.1. *Let n be a non-zero integer. A mollifier is a smooth function $\phi:\mathbb{R}^n\to[0,\infty)$ with compact support on $\{x:\in\mathbb{R}^n\ :\ |x|\le 1\}$ such that $\int_{\mathbb{R}^n}\phi(x)dx=1$ and $\lim_{R\to\infty}R^n\phi(Rx)=\delta(x)$ for all $x\in\mathbb{R}^n$, where $\delta(x)$ is the Dirac delta function. For $R>0$, define the function $\phi_R:\mathbb{R}^n\to[0,\infty)$ be the function $x\mapsto R^n\phi(Rx)$.*

An example of a mollifier is $\phi(x)=e^{-1/(1-|x|)}\mathbf{1}_{|x|<1}/\int_{|x|<1}e^{-1/(1-|y|)}dy$. The following lemma, which is standard, shows how a mollifier can be used to generate a smooth function from a continuous one.

Lemma 1.2. *Let $F:\mathbb{R}^n\to\mathbb{R}$ be continuous, and define the function $F_R(x):=\int_{\mathbb{R}^n}F(x-y)\phi_R(y)dy$. Then the function $F_R(x)$ is smooth and $\lim_{R\to\infty}F_R(x)=F(x)$ for all $x\in\mathbb{R}^n$.*

1.2 Assumptions

The following assumptions will hold throughout this paper.

$(\mathbf{A}_{\mathbf{b},\sigma})$ X is a solution to the stochastic differential equation (SDE)

$$X_0=x_0,\quad X_t=x_0+\int_0^t b(s,X_s)ds+\int_0^t\sigma(s,X_s)dW_s\quad t>0, \tag{1.6}$$

where $x_0\in\mathbb{R}^d$ is fixed and b and σ satisfy

(a) $(t,x)\in[0,T]\times\mathbb{R}^d\mapsto b(t,x)$ is $\mathbb{R}^d$-valued, measurable and uniformly bounded. Moreover, $b(t,\cdot)$ is twice continuously differentiable with uniformly bounded derivatives and Hölder continuous second derivative, and $b(\cdot,x)$ is 1/2-Hölder continuous uniformly in x.

(b) $(t,x)\in[0,T]\times\mathbb{R}^d\mapsto\sigma(t,x)$ is $\mathbb{R}^{d\times q}$-valued, measurable and uniformly bounded. Moreover, $\sigma(t,\cdot)$ is twice continuously differentiable with uniformly bounded derivatives and Hölder continuous second derivative, and $\sigma(\cdot,x)$ is 1/2-Hölder continuous uniformly in x.

(c) $\sigma(\cdot)$ satisfies a uniformly elliptic condition: there exists some finite $\bar\beta>0$ such that, for any $\zeta\in\mathbb{R}^d$, $\zeta^\top\sigma(t,x)\sigma(t,x)^\top\zeta\ge\bar\beta|\zeta|^2$ for all $(t,x)\in[0,T]\times\mathbb{R}^d$.

($\mathbf{A_\Phi}$) The terminal condition $\Phi : \mathbb{R}^d \to \mathbb{R}$ is a measurable function and there exists a constant $\alpha \in (0,1]$ such that $K^\alpha(\Phi) < \infty$, where

$$\left.\begin{aligned} K^\alpha(\Phi)^2 \quad &:= \mathbb{E}[|\Phi(X_T)|^2] + \sup_{0\le t<T} \tfrac{V_{t,T}(\Phi)^2}{(T-t)^\alpha} \\ \text{for } V_{t,T}(\Phi)^2 \quad &:= \mathbb{E}[|\Phi(X_T) - \mathbb{E}_t[\Phi(X_T)]|^2]. \end{aligned}\right\} \tag{1.7}$$

We say that Φ is fractionally smooth, and that it belongs to the space $\mathbf{L}_{2,\alpha}$. We refer to [GM10] for further discussion of and references for the space $\mathbf{L}_{2,\alpha}$.

($\mathbf{A_f}$) The driver $f : [0,T) \times \mathbb{R}^d \times \mathbb{R} \times (\mathbb{R}^q)^\top \to \mathbb{R}$ satisfies (1.2).

The following condition will be required for both the Euler scheme and the Malliavin weights scheme convergence results; this is a standard assumption for BSDE approximation schemes in order to obtain a convergence bounded from above by $O(N^{-1})$.

($\mathbf{A_{f_t}}$) The driver $f(t,x,y,z)$ is $\frac{1}{2}$-Hölder continuous in its t uniformly in (x,y,z) with Hölder constant L_f.

Our convergence results for the Malliavin weights scheme require stronger conditions than those of the Euler scheme; one of the following assumptions will be necessary to obtain the main result, Theorem 5.5, of Section 5.

($\mathbf{A_{exp\Phi}}$) The terminal condition has exponential bounds in the sense that there is a finite $C_\xi > 0$ such that $\mathbb{E}[e^{|\Phi(X_T)|}] \le C_\xi$.

($\mathbf{A_{h\Phi}}$) The function Φ is Hölder continuous: there exists a finite constants K_Φ and $\theta_\Phi \in (0,1]$ such that $|\Phi(x_1) - \Phi(x_2)| \le K_\Phi |x_1 - x_2|^{\theta_\Phi}$ for any $x_1, x_2 \in \mathbb{R}^d$.

The following assumptions will be needed for partial results only. They will hold only when specifically stated.

($\mathbf{A_{\partial f}}$) The driver $(t,x,y,z) \mapsto f(t,x,y,z)$ is continuously differentiable with respect (x,y,z) for all $t \in [0,T)$. The partial derivatives in (y,z) are bounded by $L_f(T-t)^{(\theta_L-1)/2}$ and the partial derivatives in x are bounded above by $L_X(T-t)^{1-\theta_X/2}$.

($\mathbf{A_{b\Phi}}$) The function Φ is uniformly bounded: $\|\Phi\|_\infty < \infty$.

Remark 1.3. Due to ($\mathbf{A_{b,\sigma}}$), ($\mathbf{A_{h\Phi}}$) implies ($\mathbf{A_{exp\Phi}}$) and ($\mathbf{A_\Phi}$). Note that it is possible that $\theta_\Phi < \alpha$: see [GGG12, page 2086, e.g. (i)].

In the proofs below, it will be necessary to compute a right-inverse to the matrix $\sigma(\cdot)$, i.e., for every $(t,x) \in [0,T) \times \mathbb{R}^d$, it will be necessary to find a (q,d)-dimensional matrix $\sigma^{-1}(t,x)$ such that $\sigma(t,x)\sigma^{-1}(t,x) = I_d$. In the case where the dimensions d and q are equal, this is uniquely defined by usual matrix inverse of $\sigma(t,x)$, whose existence is guaranteed by the uniform ellipticity condition ($\mathbf{A_{u.e.}}$). If the dimensions d and q are not equal, $\sigma^{-1}(t,x)$ is defined by the pseudoinverse $\sigma(t,x)^\top\big(\sigma(t,x)\sigma(t,x)^\top\big)^{-1}$; this is well defined because the uniform ellipticity condition ($\mathbf{A_{u.e.}}$) guarantees the existence of the inverse of $\sigma\sigma^\top$.

2 Key preliminary results

2.1 Malliavin calculus

We recall briefly some properties and definitions of Malliavin calculus. For details, we refer the reader to [Nua06].

For any $m \geq 1$, define $C_p^\infty(\mathbb{R}^m)$ to be the space of functions taking values in $\mathbb{R}$ which are infinitely differentiable such that all partial derivatives have at most polynomial growth, and denote by $W(h) := \int_0^T h_t dW_t$ the Itô integral of the $(\mathbb{R}^q)^\top$-valued, deterministic function $h \in \mathbf{L}_2([0,T];(\mathbb{R}^q)^\top)$. Let $\mathcal{R} \subset \mathbf{L}_2(\mathcal{F}_T)$ be the subspace containing all random variables F of the form $f(W(h_1),\ldots,W(h_m))$ for $h_i \in \mathbf{L}_2([0,T);\mathbb{R}^q)$ and any finite m. Define the derivative operator $D : \mathcal{R} \mapsto \mathbf{L}_2([0,T]\times\Omega)$ by $D_t F := \sum_{i=1}^m \partial_i f(W(h_1),\ldots,W(h_m))h_i(t)$. The derivative operator is extended to $\mathbb{D}^{1,2} \subset \mathbf{L}_2(\mathcal{F}_T)$, the closure of $\mathcal{R}$ in $\mathbf{L}_2(\mathcal{F}_T)$ under the norm $\|F\|_{1,2}^2 := \|F\|_2^2 + \mathbb{E}[\int_0^T |D_t F|^2 dt]$. Define by $\mathbb{D}^{1,2}(\mathbb{R}^k)$ (resp. $\mathbb{D}^{1,2}((\mathbb{R}^k)^\top)$) by the space of random variables $F = (F_1,\ldots,F_k)^\top$ (resp. $F = (F_1,\ldots,F_k)$) such that $F_i \in \mathbb{D}^{1,2}$ for each $i \in \{0,\ldots,k\}$. The Mallivin derivative DF is denoted by the $\mathbb{R}^{k\times q}$- (resp. $\mathbb{R}^{q\times k}$-) valued process whose i-th row (resp. column) is DF_i (resp. $(DF_i)^\top$).

The following lemma, termed the *chain rule* of Malliavin calculus, is proved in [Nua06, Proposition 1.2.3].

Lemma 2.1 (Chain rule)**.** *Let $(F_1,\ldots,F_m) \in (\mathbb{D}^{1,2})^m$. For any continuously differentiable function $f : \mathbb{R}^m \to \mathbb{R}$ with bounded partial derivatives, and $F = f(F_1,\ldots,F_m) \in \mathbb{D}^{1,2}$, the random variable $f(F) \in \mathbb{D}^{1,2}$ and $Df(F) = \sum_{i=1}^m \partial_i f(F) DF_i = \nabla_x f(F) DF$.*

Remark. In the case that F takes values in $(\mathbb{R}^m)^\top$, the result of Lemma 2.1 hold with $Df(F) = \nabla_x f(F)(DF)^\top$. In the case that f takes values in $(\mathbb{R}^k)^\top$, applying Lemma 2.1 component-wise yields that $f(F)$ is in $\mathbb{D}^{1,2}((\mathbb{R}^k)^\top)$ and $Df(F) = (DF)^\top \nabla_x f(F)$.

For the space

$$\mathrm{dom}(\delta) := \{u \in \mathbf{L}_2([0,T]\times\Omega;(\mathbb{R}^q)^\top) : \exists c \in \mathbb{R}\ s.t\ \forall F \in \mathbb{D}^{1,2}\ |\mathbb{E}[\int_0^T (u_s \cdot D_s F) ds]| \leq c\|F\|_2^2\ \}$$

define the Skorohod integral operator $\delta : \mathrm{dom}(\delta) \to \mathbf{L}_2(\Omega)$ as the dual operator to the Malliavin derivative in the sense that $\mathbb{E}[\int_0^T (u_s \cdot D_s F) ds] = \mathbb{E}[F\delta(u)]$. Below are the key properties of the Skorohod integral used in this paper.

Lemma 2.2 (Integration-by-parts)**.** *Suppose that $u \in dom(\delta)$ and $F \in \mathbb{D}^{1,2}$ are such that $\mathbb{E}[F^2 \int_0^T |u_s|^2 ds] < \infty$. Then, the* integration by parts formula *holds: $\int_0^T (u_s \cdot D_s F) ds = F\delta(u) - \delta(Fu)$.*

Remark 2.3. Suppose that the process u takes values in $\mathbb{R}^{q\times k}$ is such that $u_i^\top$ is in $\mathrm{dom}(\delta)$ for each $i \in \{0,\ldots,k\}$, where u_i is the i-th column of u. The Skorohod integral of u, denoted by $\delta(u)$, is defined by

$$\delta(u) := \big(\delta(u_1^\top)^\top,\ldots,\delta(u_k^\top)^\top\big). \tag{2.1}$$

The integration by parts formula, Lemma 2.2, is applied column-wise in the case of matrix valued u. where $D_s F u_s$ is understood as a matrix-matrix multiplication, and the Skorohod integrals are defined in the multidimensional sense of equation (2.1).

2.2 SDEs and Malliavin calculus

Fix $t \in [0,T)$ and $x \in \mathbb{R}^d$. We recall some standard properties on the Malliavin calculus applied to SDEs $X^{(t,x)}$ of the form

$$X_s^{(t,x)} = x + \int_t^s b(r, X_r^{(t,x)}) \mathbf{1}_{(t,T]}(r) dr + \int_t^s \sigma(r, X_r^{(t,x)}) \mathbf{1}_{(t,T]}(r) dW_r. \tag{2.2}$$

Observe that the SDE X defined in (1.6) is equal to $X^{(0,x_0)}$. First, we recall the flow $\nabla X^{(t,x)}$ and its inverse $\nabla X^{(t,x,-1)}$, which are respectively defined as the solutions to the SDEs

$$\nabla X_r^{(t,x)} = I_d + \int_t^r \nabla_x b(u, X_u^{(t,x)}) \nabla X_u^{(t,x)} du + \sum_{j=1}^q \int_0^r \nabla_x \sigma_j(u, X_u^{(t,x)}) \nabla X_u^{(t,x)} dW_{j,u},$$

$$\begin{aligned}\nabla X_r^{(t,x,-1)} = I_d &+ \int_t^r \nabla X_u^{(t,x,-1)} \big(\sum_{j=1}^q (\nabla_x \sigma_j(u, X_u^{(t,x)}))^2 - \nabla_x b(u, X_u) \big) du \\ &- \sum_{j=1}^q \int_t^r \nabla X_u^{(t,x,-1)} \nabla_x \sigma_j(u, X_u^{(t,x)}) dW_{j,u},\end{aligned}$$

where σ_j is the j-th column of σ. These processes are linear SDEs, and we list some standard properties used throughout this paper in the following Lemma.

Lemma 2.4. *For every $p > 1$, $\nabla X^{(t,x)}$ and $\nabla X^{(t,x,-1)}$ are in $\mathcal{S}^p$, and there is a constant C_p depending only on $\|\sigma\|_\infty$, $\|\nabla_x b\|_\infty$, $\|\nabla_x \sigma_j\|_\infty$, T and p such that*

$$\|\nabla X^{(t,x)}\|_{\mathcal{S}^p} + \|\nabla X^{(t,x,-1)}\|_{\mathcal{S}^p} \leq C_p.$$

Moreover,

$$\|\nabla X_r^{(t,x)} - \nabla X_s^{(t,x)}\|_2^2 + \|\nabla X_r^{(t,x,-1)} - \nabla X_s^{(t,x,-1)}\|_2^2 \leq C_2 |r-s|$$

for all $(t,s) \in [0,T]^2$,

$$\nabla X_r^{(t,x)} \nabla X_r^{(t,x,-1)} = I_d$$

for all $r \in [t,T]$ almost surely, and, for any $r < u < s$,

$$\mathbb{E}_r[|\nabla X_s^{(t,x)} \nabla X_r^{(t,x,-1)} - \nabla X_u^{(t,x)} \nabla X_r^{(t,x,-1)}|^2] \leq C_2(s-u) \quad \mathbb{P}-a.s.$$

The Malliavin derivative of the marginals of $X^{(t,x)}$ is strongly related to the flow and its inverse, as shown in the following Lemma. The proof of the estimates follows directly from Lemma 2.4.

Lemma 2.5. *For all $r \in [0,T]$, $X_r^{(t,x)}$ is in $\mathbb{D}^{1,2}(\mathbb{R}^d)$ and there is a version $D_s X_r^{(t,x)}$ satisfying the SDE*

$$D_s X_r^{(t,x)} \mathbf{1}_{[s,T]}(r) = \Big\{ \sigma(s, X_s^{(t,x)}) + \int_s^r \nabla_x b(\tau, X_\tau^{(t,x)}) D_s X_\tau^{(t,x)} d\tau + \sum_{j=1}^q \int_s^r \nabla_x \sigma_j(\tau, X_\tau^{(t,x)}) D_s X_r^{(t,x)} dW_{j,r} \Big\}.$$

Moreover, for all $0 \leq s, r \leq T$,

$$D_s X_r^{(t,x)} = \nabla X_r^{(t,x)} \nabla X_s^{(t,x,-1)} \sigma(s, X_s^{(t,x)}) \mathbf{1}_{[s,T]}(r) \mathbf{1}_{[t,T]}(s) \quad a.s.$$

whence there exists a constant C_p depending only on $\|\sigma\|_\infty$, $\|\nabla_x b\|_\infty$, $\|\nabla_x \sigma_j\|_\infty$, T and p such that $\mathbb{E}_s[|D_s X_r^{(t,x)}|^p] \leq C_p$, and $\sup_s \mathbb{E}[\sup_{s \leq r \leq T} |D_s X_r^{(t,x)}|^2]^{1/2} \leq C_2$; moreover, for any $x_1, x_2 \in \mathbb{R}^d$, $\mathbb{E}[|D_s X_r^{(t,x_1)} - D_s X_r^{(t,x_2)}|^p] \leq C_p |x_1 - x_2|^p$ and, for any $r < u < s$, $\mathbb{E}_r[|D_r X_s^{(t,x)} - D_r X_u^{(t,x)}|^2] \leq C_2(s-u)$.

2.3 Existence, uniqueness, approximation and decomposition of the BSDE

Since the class of BSDEs under local conditions has, to the best of our knowledge, not been studied in full generality, we now include a proof of the existence and uniqueness of solutions. We remark that the existence and uniqueness follows also from [FJ12, Theorem 3.2]. The proof below is simpler, since a simpler class of BSDEs is considered, and different, so we include for the interest of the reader.

Theorem 2.6. *There exists a unique pair of process* (Y,Z) *in* $\mathcal{S}^2\times\mathcal{H}^2$ *solving the BSDE* (1.1) *with terminal condition* $\Phi(X_T)\in\mathbf{L}_2(\mathcal{F}_T)$ *and driver* f *satisfiying the locally Lipschitz continuous and boundedness of* (1.2).

Proof. Let (ϕ,ψ) be in $\mathcal{H}^2\times\mathcal{H}^2$, and define the random function $f(r,y,z)=f(r):=f(r,X_r,\phi_r,\psi_r)$. We show that there exists a unique solution $(Y^{(\phi,\psi)},Z^{(\phi,\psi)})$ to the BSDE

$$Y_t^{(\phi,\psi)}=\Phi(X_T)+\int_t^T f(r)dr-\sum_{j=1}^q\int_t^T Z_{j,r}^{(\phi,\psi)}dW_{j,r}.$$

in $\mathcal{H}^2\times\mathcal{H}^2$ (in fact, $Y^{(\phi,\psi)}$ is in $\mathcal{S}^2$). This will imply the function $\Xi:\mathcal{H}^2\times\mathcal{H}^2\to\mathcal{H}^2\times\mathcal{H}^2$ mapping (ϕ,ψ) to $(Y^{(\phi,\psi)},Z^{(\phi,\psi)})$ is well defined. For this, we use [BDH+03, Theorem 4.2]. The function f is predictably measurable; we must show that f satisfies assumptions (H1)-(H5) of [BDH+03, Section 4]. Since f takes no argument in (y,z), it is only necessary to check (H1), which follows readily the local Lipschitz continuity and local boundedness of the driver (1.2). Therefore, from [BDH+03, Theorem 4.2], $(Y^{(\phi,\psi)},Z^{(\phi,\psi)})$ exists and is unique. As in the proof of [EKPQ97, Theorem 2.1], we prove that Ξ is a contraction. For $k\in\{1,2\}$, let $(\phi_k,\psi_k)\in\mathcal{H}^2\times\mathcal{H}^2$ and define the BSDE $(Y_k,Z_k):=\Xi(\phi_k,\psi_k)$. Define the differences $\delta Y=Y_1-Y_2$, $\delta Z=Z_1-Z_2$, $\delta\phi=\phi_1-\phi_2$ and $\delta\psi=\psi_1-\psi_2$. It then follows from Hölder's inequality that

$$\begin{aligned}\|\delta Y_t\|_2^2+\int_t^T\|\delta Z_r\|_2^2dr&\le\|\int_t^T|f(r,X_r,\phi_{1,r},\psi_{1,r})-f(r,X_r,\phi_{2,r},\psi_{2,r})|dr\|_2^2\\&\le L_f^2(T-t)^{\theta_L}\int_t^T\{\|\delta\phi_r\|_2^2+\|\delta\psi_r\|_2^2\}dr\end{aligned}$$

for all $t\in[0,T)$. Setting $t_0=(T-1/(4L_f^2)^{1/\theta_L}\wedge 1)\vee 0$ ensures, on the one hand, that $L_f^2(T-t_0)^{\theta_L}\le 1/4$, and, on the other hand, that $T-t_0\le 1$. Integrating the above inequality on the interval $t\in[t_0,T)$ then yields $4\int_{t_0}^T\{\|\delta Y_r\|_2^2+\|\delta Z_r\|_2^2\}dr\le\int_{t_0}^T\{\|\delta\phi_r\|_2^2+\|\delta\psi_r\|_2^2\}dr$ and $4\|\delta Y_t\|_2^2\le\int_{t_0}^T\{\|\delta\phi_r\|_2^2+\|\delta\psi_r\|_2^2\}dr$ for all $t\in[t_0,T)$. On the interval $[0,t_0)$, the function $f(t,x,\cdot)$ is Lipschitz continuous with a uniform Lipschitz constant for all (t,x), so we proceed as in the proof of Theorem [EKPQ97, Theorem 2.1] to show that, for sufficiently large $\eta>0$,

$$\int_0^{t_0}e^{\eta r}\{\|\delta Y_r\|_2^2+\|\delta Z_r\|_2^2\}dr\le e^{\eta t_0}\|\delta Y_{t_0}\|_2^2+\frac{1}{2}\int_0^{t_0}e^{\eta r}\{\|\delta\phi_r\|_2^2+\|\delta\psi_r\|_2^2\}dr$$

Combining this with the above estimates on $\int_{t_0}^T\{\|\delta Y_r\|_2^2+\|\delta Z_r\|_2^2\}dr$ and $\|\delta Y_{t_0}\|_2^2$ then yields

$$\int_0^T e^{\eta_r}\{\|\delta Y_r\|_2^2+\|\delta Z_r\|_2^2\}dr\le\frac{1}{2}\int_0^T e^{\eta_r}\{\|\delta\phi_r\|_2^2+\|\delta\psi_r\|_2^2\}dr$$

where $\eta_r=\eta(r\wedge t_0)$. This is sufficient to prove that Ξ is a contraction. □

We now introduce an approximation procedure that will be used repeatedly in this paper; we introduce intermediate BSDEs by "cutting" the tail of the driver close to the time horizon T, prove our results for these BSDEs, then extend the result to the BSDE we're interested by limiting procedures. This technique was used extensively in [GM10], and we shall frequently take advantage of it throughout this work.

Definition 2.7. *Let $(t, \varepsilon) \in [0, T)^2$ and define $f^{(\varepsilon)}(t, x, y, z) := f(t, x, y, z)\mathbf{1}_{[0,T-\varepsilon)}(t)$. Let $(Y^{(\varepsilon)}, Z^{(\varepsilon)})$ be the solution of the BSDE*

$$Y_t^{(\varepsilon)} = \Phi(X_T) + \int_t^T f^{(\varepsilon)}(s, X_s, Y_s^{(\varepsilon)}, Z_s^{(\varepsilon)})ds - \int_t^T Z_s^{(\varepsilon)} dW_s. \tag{2.3}$$

Additionally, let (y, z) be the solution of the BSDE with zero driver $y_t = \Phi(X_T) - \int_t^T z_s dW_s$ and $(y^{(\varepsilon)}, z^{(\varepsilon)})$ the solution of the BSDE with zero terminal condition

$$y_t^{(\varepsilon)} = \int_t^T f^{(\varepsilon)}(s, X_s, y_s + y_s^{(\varepsilon)}, z_s + z_s^{(\varepsilon)})ds - \int_t^T z_s^{(\varepsilon)} dW_s. \tag{2.4}$$

Since $f^{(\varepsilon)}(t, x, y, z)$ is Lipschitz continuous uniformly in t with Lipschitz constant $L_f \varepsilon^{(\theta_L - 1)/2}$, the solutions of the BSDEs in Definition 2.7 exists in $\mathcal{S}^2 \times \mathcal{H}^2$ and are unique for all $\varepsilon \in [0, T)$ [EKPQ97, Theorem 2.1]. We shall also make use of the decomposition $(Y^{(\varepsilon)}, Z^{(\varepsilon)}) = (y + y^{(\varepsilon)}, z + z^{(\varepsilon)})$, which is standard in BSDE literature [GM10].

We first treat the linear BSDE (y, z). The following Lemma relates the linear BSDE (y, z) to the PDE in (2.5) and gives some boundedness properties for the function u and its derivatives; these bounds will be used throughout this paper.

Lemma 2.8. *Let $(\mathbf{A_{b\Phi}})$ be in force and consider the PDE*

$$\left.\begin{aligned} &0 = \partial_t u + \tfrac{1}{2}\textstyle\sum_{i,j=1}^d (\sigma\sigma^\top)_{i,j} \frac{\partial^2}{\partial_{x_i}\partial_{x_j}} u + \sum_{i=1}^d b_i \frac{\partial}{\partial_{x_i}} u, \\ &u(T, x) = \Phi(x). \end{aligned}\right\} \tag{2.5}$$

Then, for all $(t, x) \in [0, T] \times \mathbb{R}^d$,

$$u(t, x) = \mathbb{E}[\Phi(X_T)|X_t = x]$$

is a classical solution of the PDE (2.5) (the so-called Feynman-Kac representation). The derivatives $\partial_x^\alpha u$ ($|\alpha| \leq 3$), $\partial_t u$, $\partial_t \nabla_x u$ exist and are continuous. There is a constant C depending only on the bound on b and it's derivatives, the bound on σ and it's derivatives, and $\bar{\beta}$ such that

$$\|\partial_x^\alpha u(t, \cdot)\|_\infty \leq C\|\Phi\|_\infty (T - t)^{-|\alpha|/2}$$

for all $(t, x) \in [0, T) \times \mathbb{R}^d$. Moreover, $\big(u(t, X_t), (\nabla_x u(t, X_t)\sigma(t, X_t))^\top\big)$ is the solution to the linear BSDE (y, z). For any $x_1, x_2 \in \mathbb{R}^d$, $t \in [0, T)$ recall from (2.2) the SDEs $X^{(t,x_1)}$ and $X^{(t,x_2)}$, and for $\alpha, \beta \in [0, 1]$ define $\bar{X} := \alpha X^{(t,x_1)} + \beta X^{(t,x_2)}$; then

$$|\nabla_x u(r, \bar{X}_r)|^2 \leq \frac{C\mathbb{E}_r[|\Phi(\bar{X}_T) - \mathbb{E}_r[\Phi(\bar{X}_T)]|^2]}{(T - r)} \quad \text{and} \quad |\nabla_x^2 u(r, \bar{X}_r)|^2 \leq \frac{C\mathbb{E}_r[|\Phi(\bar{X}_T) - \mathbb{E}_r[\Phi(\bar{X}_T)]|^2]}{(T - r)^2}$$

for all $r \in [0, T)$.

Proof. The Feynman-Kac representation of the solution is well known, see [GM$^+$05] among others. To obtain the gradient bounds, recall that X is a Markov process and denote its transition density by $p(t, x; s, \xi)$. For some C_1 and β finite, the following gradient bounds hold on $p(t, x; s, \xi)$:

$$|\partial_x^\alpha p(t, x; s, \xi)| \leq \frac{C_1 e^{\beta|x-\xi|^2/(s-t)}}{(s - t)^{(d+|\alpha|)/2}} \quad \text{for } |\alpha| \leq 3,$$

$$|\partial_t \partial_x^\alpha p(t, x; s, \xi)| \leq \frac{C_1 e^{\beta|x-\xi|^2/(s-t)}}{(s - t)^{(d+2+|\alpha|)/2}} \quad \text{for } |\alpha| \leq 1.$$

We obtained these bounds from [GL10, Appendix A], who provide references for proofs. The bounds on the derivatives of $u(t,\cdot)$ then follow from Lebesgue's differentiation theorem (differentiation with respect to t and x) applied to

$$\partial_t^{\alpha_0}\partial_x^{\alpha} u(t,x) = \partial_t^{\alpha_0}\partial_x^{\alpha}\int_{\mathbb{R}^d}\Phi(\xi)p(t,x;T,\xi)d\xi = \int_{\mathbb{R}^d}\Phi(\xi)\partial_t^{\alpha_0}\partial_x^{\alpha}p(t,x;T,\xi)d\xi$$

for multiindices α_0 and α; we apply the gradient bounds on the transition density above and the boundedness of Φ to obtain the result on $|\partial_t^{\alpha_0}\partial_x^{\alpha}u(t,x)|$.

To show the bound on $|\nabla_x u(r,\bar{X}_r)|$, let us recall first that the result in the case $\alpha = 1$ and $\beta = 0$ is given in [GM10, Lemma 1.1]. The authors use the tools of [GM$^+$05, Lemma 2.9] to show that, for every $r \in [0,T)$ and $x \in \mathbb{R}^d$, there is a $\mathcal{F}_T$-measurable random variable $H_{r,x}$ such that

$$\nabla_x u(r, X_r^{(t,x)}) = \mathbb{E}_r[(\Phi(X_T) - \mathbb{E}_r[\Phi(X_T)])H_{r,x}].$$

This result follows largely from the integration-by-parts formula of Malliavin calculus – Lemma 2.2 – and martingale arguments; see the proof of [GM$^+$05, Lemma 2.9] for details. $H_{r,x}$ satisfies $\mathbb{E}_r[H_{r,x} = 0]$ and $\mathbb{E}_r[|H_{r,x}|^2] \leq C(T-r)^{-1}$. The result for $(\alpha,\beta) = (1,0)$ then follows by the Cauchy-Schwarz inequality. (Note that we in fact don't need $(\mathbf{A_{b\Phi}})$ to obtain this result.) One can follow the proof method of [GM$^+$05, Lemma 2.9], using additionally the linearity of the Malliavin derivative, to show that

$$\nabla_x u(r,\bar{X}_r) = \mathbb{E}_r[\Phi(\bar{X}_T)\bar{H}_r],$$

where $\bar{H}_r := \alpha H_{r,x_1} + \beta H_{r,x_2}$, whence the result follows. The proof for the bound on $|\nabla_x^2 u(r,\bar{X}_r)|$ is similar. □

We move onto the non-linear BSDE $(y^{(\varepsilon)}, z^{(\varepsilon)})$. The following representations and a priori estimates will be critical throughout this paper.

Lemma 2.9. *Let $(\mathbf{A_{\partial f}})$ and $(\mathbf{A_{b\Phi}})$ hold. Recall the function $u:[0,T]\times\mathbb{R}^d \to \mathbb{R}$ solving the PDE (2.5) and that it is differentiable (Lemma 2.8), define $\Theta_r = (r, X_r, Y_r^{(\varepsilon)}, Z_r^{(\varepsilon)})$, and set*

$$\begin{aligned} a_r^{(\varepsilon)} &:= \nabla_x f^{(\varepsilon)}(\Theta_r) + \nabla_y f^{(\varepsilon)}(\Theta_r)\nabla_x u(r,X_r) + \nabla_z f^{(\varepsilon)}(\Theta_r)U(r,X_r)^\top, \\ b_r^{(\varepsilon)} &:= \nabla_y f^{(\varepsilon)}(r,X_r,Y_r^{(\varepsilon)},Z_r^{(\varepsilon)}), \quad c_r^{(\varepsilon)} := \nabla_z f^{(\varepsilon)}(r,X_r,Y_r^{(\varepsilon)},Z_r^{(\varepsilon)}) \end{aligned} \tag{2.6}$$

where the gradients $\nabla_\xi f(\Theta_r)$ is given by $\nabla_\xi f(r,x,y,z)|_{(r,x,y,z)=\Theta_r}$ for $\nabla_\xi f(r,x,y,z)$ defined as in Section 1.1, and $U(r,x)$ is defined by

$$U(t,x) := \nabla_x^2 u(t,x)\sigma(t,x) + \sum_{j=1}^d (\nabla_x u)_j(t,x)\nabla_x\sigma_j^\top(t,x), \tag{2.7}$$

Then there a finite constant C depending only on T, d, $K^\alpha(\Phi)$, the bounds on b and σ and their derivatives, L_f, and θ_L such that

$$\|a_r^{(\varepsilon)}\|_2 \leq C\mathbf{1}_{[0,T-\varepsilon)}(r)(T-r)^{(\alpha+\theta_L-3)/2} \tag{2.8}$$

There exists a unique solution $(U^{(\varepsilon)}, V^{(\varepsilon)}) \in \mathcal{S}^2 \times \mathcal{H}^2$ of the BSDE

$$\begin{aligned} U_t^{(\varepsilon)} = \int_t^T a_r^{(\varepsilon)} + U_r^{(\varepsilon)}\big\{b_r^{(\varepsilon)} I_d + \nabla_x b(r,X_r) + \sum_{j=1}^q c_{j,r}^{(\varepsilon)}\nabla_x\sigma_j(r,X_r)\big\}dr \\ + \int_t^T \sum_{j=1}^q (V_{j,r}^{(\varepsilon)})^\top\big\{c_{j,r}^{(\varepsilon)} I_d + \nabla_x\sigma_j(r,X_r)\big\}dr - \sum_{j=1}^q\int_t^T (V_{j,r}^{(\varepsilon)})^\top dW_{j,r} \end{aligned} \tag{2.9}$$

where $\sigma_j(\cdot)$ is the j-th column of $\sigma(\cdot)$, $c^{(\varepsilon)}_{r,j}$ is the j-th component of $c^{(\varepsilon)}_r$, and $V^{(\varepsilon)}_{j,r}$ is the j-th column of $V^{(\varepsilon)}_r$. There is a (possibly different) constant C such that, for any $0 \le t < T$ and $\varepsilon > 0$,

$$\mathbb{E}[\sup_{t\le r<T} |U^{(\varepsilon)}_r|^2] + \int_t^T \|V^{(\varepsilon)}_r\|_2^2 dr \le C\| \int_t^{T-\varepsilon} |a^{(\varepsilon)}_r| dr\|_2^2 \le \frac{C}{\varepsilon^{1-(\theta_L+\alpha)\wedge 1}}. \tag{2.10}$$

Let us consider $(\nabla y^{(\varepsilon)}, \nabla z^{(\varepsilon)})$ solving the BSDE

$$\begin{aligned}\nabla y^{(\varepsilon)}_t &= \int_t^T \nabla_x f^{(\varepsilon)}(\Theta_r)\nabla X_r + \nabla_y f^{(\varepsilon)}(\Theta_r)\{\nabla_x u(r, X_r)\nabla X_r + \nabla y^{(\varepsilon)}_r\} dr \\ &\quad + \int_t^T \nabla_z f^{(\varepsilon)}(\Theta_r) U(r, X_r)^\top \nabla X_r + \sum_{j=1}^q \nabla_z f^{(\varepsilon)}_j(\Theta_r)(\nabla z^{(\varepsilon)}_{j,r})^\top dr - \sum_{j=1}^q \int_t^T (\nabla z^{(\varepsilon)}_{j,r})^\top dW_r.\end{aligned} \tag{2.11}$$

The processes $z^{(\varepsilon)}$ and $\nabla z^{(\varepsilon)}$ satisfy the representations

$$z^{(\varepsilon)}_t = U^{(\varepsilon)}_t \sigma(t, X_t) \quad m \times \mathbb{P} - a.e. \tag{2.12}$$

$$(V^{(\varepsilon)}_{j,t})^\top = (\nabla z^{(\varepsilon)}_{j,t})^\top \sigma^{-1}(t, X_t) - U^{(\varepsilon)}_t \nabla_x \sigma_j(t, X_t)\; m \times \mathbb{P} - a.e. \tag{2.13}$$

where $\nabla z^{(\varepsilon)}_{j,t}$ is the j-th column of $\nabla z^{(\varepsilon)}_t$.

Proof. In what follows, C may change from line to line. From [GM10, Lemma 1.1], $\|\nabla_x u(t, X_t)\|_2 \le C(T-t)^{(\alpha-1)/2}$ and $\|\nabla_x^2 u(t, X_t)\|_2 \le C(T-t)^{(\alpha-2)/2}$. Therefore, $\|a^{(\varepsilon)}_r\|_2 \le C(T-r)^{(\theta_L+\alpha-3)/2}$ for all $r \in [0, T-\varepsilon]$, which is the bound (2.8), whence

$$(\int_0^{T-\varepsilon} \|a^{(\varepsilon)}_r\|_2 dr)^2 < \frac{C}{\varepsilon^{1-(\theta_L+\alpha)\wedge 1}} < \infty.$$

This is the second inequality in (2.10). Additionally, for all $t \in [0, T)$, $|b^{(\varepsilon)}_t| + \max_j |c^{(\varepsilon)}_{j,t}| \le C(T-t)^{(\theta_L-1)/2}$ almost surely. The first inequality in (2.10) follows. Let (ϕ, ψ) be a $(\mathbb{R}^d)^\top \times \mathbb{R}^{d\times q}-$ valued process in $\mathcal{H}^2$, and define the random function

$$\begin{aligned}g(r, y, z) = g(r) &:= a^{(\varepsilon)}_r + \phi_r((b^{(\varepsilon)}_r I_d + \nabla_x b(r, X_r) + \sum_{j=1}^q c^{(\varepsilon)}_{j,r} \nabla_x \sigma_j(r, X_r)) \\ &\quad + \sum_{j=1}^q (\psi_{j,r})^\top (c^{(\varepsilon)}_{j,r} I_d + \nabla_x \sigma_j(r, X_r)).\end{aligned}$$

The function g is progressively measurable and satisfies assumptions (H1)-(H5) of [BDH+03, Section 4]. Since f takes no argument in (y, z), it is only necessary to validate (H1): using the triangle inequality, Jensen's inequality, the Cauchy-Schwarz inequality, and assumptions $(\mathbf{A}_{\partial\mathbf{f}})$ and $(\mathbf{A}_{\mathbf{b},\sigma})$, it follows that

$$\begin{aligned}\mathbb{E}[(\int_0^T |g(r)| dr)^2]^{1/2} &\le \mathbb{E}[(\int_0^T |a^{(\varepsilon)}_r| dr)^2]^{1/2} + C(\int_0^T \mathbb{E}[|\phi_r|^2] dr)^{1/2} (\int_0^T \frac{dr}{(T-r)^{1-\theta}})^{1/2} \\ &\quad + C\sum_{j=1}^q (\int_0^T \mathbb{E}[|\psi_{j,r}|^2] dr)^{1/2} (\int_0^T \frac{dr}{(T-r)^{1-\theta}})^{1/2} < \infty.\end{aligned}$$

Thanks to [BDH+03, Theorem 4.2], there exists a unique solution (u, v) to the BSDE

$$u_t = \int_t^T g(r) dr - \sum_{j=1}^q \int_t^T v_{j,r} dW^j_r \qquad t \in [0, T).$$

in $\mathcal{S}^2 \times \mathcal{H}^2$. The remainder of the proof of existence and uniqueness follows exactly as the proof of Theorem 2.6. To prove the first inequality in (2.10), observe that the driver $g(r)$ satisfies (A.1) from Proposition A.2 with $f_r = |a_r^{(\varepsilon)}|$ and $\lambda_r = \mu_r = C(T-r)^{(\theta-1)/2}$.

The proofs of (2.12) and (2.13) are given in [GM10, Theorem 2.1]. The inclussion of the local Lipschitz continuity assumptions (1.2) make no difference, because the driver $f(t,x,y,z)\mathbf{1}_{[0,T-\varepsilon)}(t)$ is Lipschitz continuous uniformly in t in (x,y,z) with Lipschitz coefficient $L_f\varepsilon^{(\theta_L-1)/2}$. $\square$

2.4 A priori estimates

For $0 \le s < r \le T$, we define the Malliavin weights by

$$H_r^s := \frac{1}{r-s}\Big(\int_s^r (\sigma^{-1}(t,X_t)D_sX_t)^\top dW_t\Big)^\top \tag{2.14}$$

where D_sX_t is the Malliavin derivative of X_t at s defined in Section 2.2. It was shown in Lemma 4.1 that $|\sigma^{-1}(t,x)|$ is uniformly bounded in (t,x). The following constant appears throughout this paper

$$C_M := \|\sigma^{-1}\|_\infty^2 \sup_{s\in[0,T)} \sup_{t\in(s,T]} \mathbb{E}_s[|D_sX_t|^2]. \tag{2.15}$$

It is known from Lemma 2.4 that $\sup_{s\in[0,T)}\sup_{t\in(s,T]}\mathbb{E}_s[|D_sX_t|^2]$ is bounded. The following result is used in the proof of [GM10, Lemma 1.1]; we include it here for completeness.

Lemma 2.10. *For any $0 \le s \le r \le T$,*

$$\mathbb{E}_s[|H_r^s|^2] \le \frac{C_M}{r-s} \qquad \text{almost surely.}$$

Moreover, for every $p \ge 2$, there is a finite $C_p \ge 0$ depending only on p, $\|\sigma\|_\infty$, $\|\nabla_x b\|_\infty$, $\max_j \|\nabla_x\sigma_j\|_\infty$, and T such that $\|H_r^s\|_p \le C_p(r-s)^{-p/2}$.

Proof. Observe, using Lemma 2.5 and the fact that $(s-r)^2|H_r^s|^2 - \int_s^r |\sigma^{-1}(t,X_t)D_sX_t|^2dt$ is a (local) martingale, that

$$\mathbb{E}_s[|H_r^s|^2] = (r-s)^{-2}\mathbb{E}_s\Big[\int_s^r |\sigma^{-1}(t,X_t)D_sX_t|^2dt\Big] \le \frac{\|\sigma^{-1}\|_\infty}{(r-s)^2}\mathbb{E}_s\Big[\int_s^r |D_sX_t|^2dt\Big].$$

One then applies the conditional Fubini's lemma, Lemma A.1, and the uniform bound on $\mathbb{E}_s[|D_sX_t|^2]$ from Lemma 2.5 to complete the proof. The bound on $\|H_r^s\|_p$ is proved using the Burkholder-Davis-Gundy inequality on the continuous local martingale $(t-s)H_t^s$. $\square$

The Malliavin weight is a critical element of this work. We use it to obtain a priori estimates in this section, to obtain the representation theorem in Section 2.5, and for the Malliavin weights scheme of Section 5. The following elementary corollary indicates an important technique in which we make use of the Cauchy-Schwarz inequality in conditional form in order to obtain upper bounds:

Corollary 2.11. *Let $G \in \mathbf{L}_2(\mathcal{F}_T)$ and $g \in \mathbf{L}_2([0,T]\times\Omega)$. Then*

$$|\mathbb{E}_t[GH_T^t]| \le \frac{\sqrt{C_M}(\mathbb{E}_t[|G|^2])^{1/2}}{\sqrt{T-t}} \quad \text{and} \quad |\mathbb{E}_t[\int_t^T \{g_sH_s^t\}ds]| \le \sqrt{C_M}\int_t^T \frac{\mathfrak{g}_s}{\sqrt{s-t}}ds$$

where $\mathfrak{g} \in \mathbf{L}_2([0,T]\times\Omega)$ is a version of $\big((\mathbb{E}_t[|g_s|^2])^{1/2}\big)_{s\in[0,T)}$.

Remark. We leave the implementation of the conditional Fubini theorem, Lemma A.1, in its full form in the above lemma, without using the notation given in Section 1.1. We do this to be absolutely clear about how the conditional Fubini theorem is used in this paper, before returning to the – in our opinion – much more clear, if slightly abusive, notation $\int_t^T(\mathbb{E}_t[g_s])^{1/2}ds$.

Proof. The first inequality follows from application of the conditional Cauchy-Schwarz inequality $|\mathbb{E}_t[GH_T^t]| \leq (\mathbb{E}_t[|G|^2])^{1/2}(\mathbb{E}_t[|H_T^t|])^{1/2}$, then using Lemma 2.10 to upper bound the conditional expectation $(\mathbb{E}_t[|H_T^t|])^{1/2}$. The second inequality is a little more intricate to obtain due to the Lebesgue integral. First, apply the conditional Fubini theorem, Lemma A.1, to obtain

$$\mathbb{E}_t[\int_t^T |g_s H_s^t| ds] = \int_t^T \mathfrak{H}_s ds$$

where $\mathfrak{H} \in \mathbf{L}_2([0,T] \times \Omega)$ is a version of $(\mathbb{E}_t[|g_s H_s^t|])_{s\in[0,T)}$. Now, applying the conditional Cauchy-Schwarz inequality and Lemma 2.10 to $\mathbb{E}_t[|g_s H_s^t|]$, it follows that

$$\mathfrak{H}_s \leq C_M \frac{\mathfrak{g}_s}{\sqrt{s-t}} \quad \text{for almost all } s \in [0,T) \quad \mathbb{P}-\text{a.s.}$$

as required. □

We now state and prove a priori results on the solutions of BSDEs with drivers satisfying (1.2). These estimates are in the spirit of [EKPQ97, Proposition 2.1] with two extensions: firstly, we allow the drivers of the BSDEs to satisfy locally Lipschitz continuity like condition ($\mathbf{A_f}$); secondly, we prove point-wise (in time) a priori estimates on the Z processes assuming the existence of a representation formula. The latter estimates will be extremely useful, as we shall prove the this representation formula for our BSDEs in Section 2.5 and use the below proposition extensively in subsequent sections.

Proposition 2.12. *Let $x \mapsto \Phi_1, \Phi_2 \in \mathbf{L}_2(\mathcal{F}_T)$and $(\omega,t,y,z) \mapsto f_1(\omega,t,y,z), f_2(\omega,t,y,z)$ be $\mathcal{P} \otimes \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}((\mathbb{R}^q)^\top)$-measurable functions for which there are constants $(\theta_{1,L},\theta_{2,L}) \in (0,1]^2$ and $(L_{f_1}, L_{f_2}) \in (0,\infty)^2$ such that*

$$|f_i(\omega,t,y,z) - f_i(\omega,t,y',z')| \leq \frac{L_{f_i}\{|y-y'| + |z-z'|\}}{(T-t)^{(1-\theta_{i,L})/2}} \quad m \times \mathbb{P} - \textit{almost everywhere},$$

and $f_i(\omega,t,0,0) \in \mathcal{H}^2$ for $i \in \{1,2\}$. Let (Y_i, Z_i) be a solution to the FBSDE with terminal condition Φ_i and driver $f_i(t,y,z)$ ($i = 1,2$ respectively).

Define

$$\begin{aligned}
&\Delta Y_t := Y_{1,t} - Y_{2,t}, \qquad \Delta Z_t := Z_{1,t} - Z_{2,t},\\
&\Delta f_t := f_1(t, Y_{1,t}, Z_{1,t}) - f_2(t, Y_{1,t}, Z_{1,t}), \quad \Delta\Phi := \Phi_1 - \Phi_2.
\end{aligned}$$

Then there is a finite constant $C \geq 0$ depending only on T, L_{f_2} and $\theta_{2,L}$ such that, for all $s < t < T$,

$$\mathbb{E}_s[\Delta Y_t^2] + \mathbb{E}_s[\int_t^T |\Delta Z_s|^2 ds] \leq C\mathbb{E}_s[\Delta\Phi^2] + C\Big(\int_t^T \mathbb{E}_s[\Delta f_r^2]^{1/2} dr\Big)^2 \tag{2.16}$$

Moreover, suppose that $Z_{i,t} := \mathbb{E}_t[\Phi_i(X_T)H_T^t + \int_t^T f_i(r, X_r, Y_{i,r}, Z_{i,r})H_r^t dr]$ for all $t \in [0,T)$ almost surely ($i = 1,2$). Then there is a (possibly different) finite constant $C \geq 0$ depending only on T, C_M, L_{f_2}, and $\theta_{2,L}$ such that,

$$(\mathbb{E}_s[|\Delta Z_t|^2])^{1/2} \leq C\frac{(\mathbb{E}_s\big[(\Delta\Phi - \mathbb{E}_t\Delta\Phi)^2\big])^{1/2}}{\sqrt{T-t}} + C\int_t^T \frac{(\mathbb{E}_s[\Delta f_r^2])^{1/2}}{\sqrt{r-t}} dr + C(\mathbb{E}_s[\Delta\Phi^2])^{1/2}(T-t)^{\theta_L/2} \tag{2.17}$$

for all $t \in [0,T)$ almost surely.

Proof. In what follows, C may change from line to line. We start by proving the result for $s = 0$; the general case is proved analogously, the only difference is that one must use the conditional version of the Minkowski, Cauchy-Schwarz (Corollary 2.11), and Hölder inequalities in the place of the usual version of these with the regular expectation. Using the definition of the BSDE (1.1),

$$\Delta Y_t + \int_t^T \Delta Z_s dW_s = \Delta\Phi + \int_t^T \Delta f_s ds + \int_t^T f_2(s, Y_{1,s}, Z_{1,s}) - f_2(s, Y_{2,s}, Z_{2,s}) ds.$$

Using (1.2) and Hölder's inequality,

$$\begin{aligned}
\|\Delta Y_t\|_2^2 + \int_t^T \|\Delta Z_s\|_2^2 ds &\le 3\|\Delta\Phi\|_2^2 + 3\|\int_t^T \Delta f_s ds\|_2^2 + 3\|\int_t^T f_2(s, Y_{1,s}, Z_{1,s}) - f_2(s, Y_{2,s}, Z_{2,s}) ds\|_2^2 \\
&\le 3\|\Delta\Phi\|_2^2 + 3\|\int_t^T |\Delta f_s| ds\|_2^2 + 3L_{f_2}^2 \|\int_t^T \frac{|\Delta Y_s| + |\Delta Z_s|}{(T-s)^{(1-\theta_{2,L})/2}} ds\|_2^2 \\
&\le 3\|\Delta\Phi\|_2^2 + 3\|\int_t^T |\Delta f_s| ds\|_2^2 + 3L_{f_2}^2 \int_t^T \frac{1}{(T-s)^{1-\theta_{2,L}}} ds \int_t^T \{\|\Delta Y_s\|_2^2 + \|\Delta Z_s\|_2^2\} ds \\
&\le 3\|\Delta\Phi\|_2^2 + 3\|\int_t^T |\Delta f_s| ds\|_2^2 + 3L_{f_2}^2 (T-t)^{\theta_{2,L}} \int_t^T \{\|\Delta Y_s\|_2^2 + \|\Delta Z_s\|_2^2\} ds \quad (2.18)
\end{aligned}$$

Setting $t_0 = (T - 1/(6L_{f_2}^2)^{1/\theta_{2,L}}) \vee 0$ ensures that $3L_{f_2}(T - t_0)^{\theta_{2,L}} \le 1/2$, and, on the other hand, that $T - t_0 \le 1$. Integrating (2.18) over (t_0, T), we obtain

$$\int_{t_0}^T \|\Delta Y_t\|_2^2 + \|\Delta Z_s\|_2^2 ds \le 6\|\Delta\Phi\|_2^2 + 6L_{f_2}^2 \|\int_{t_0}^T |\Delta f_s| ds\|_2^2 \quad (2.19)$$

Substituting (2.19) into (2.18) then yields

$$\sup_{t_0 \le t < T} \|\Delta Y_t\|_2^2 \le 6\|\Delta\Phi\|_2^2 + 6\|\int_{t_0}^T |\Delta f_s| ds\|_2^2$$

and this gives the result in the interval $[t_0, T]$.

In the interval $[0, t_0)$, the function $(y, z) \mapsto f_2(\omega, t, y, z)$ is $m \times \mathbb{P}$ Lipschitz continuous with Lipschitz constant $\tilde{L} := L_f(T - t_0)^{(\theta_{2,L}-1)/2}$. It then follows from [EKPQ97, Proposition 2.1] that

$$\sup_{0 \le t < t_0} \|\Delta Y_t\|_2^2 + \int_{t_0}^T \|\Delta Z_s\|_2^2 ds \le C\|\Delta Y_{t_0}\|_2^2 + C\|\int_0^{t_0} |\Delta f_s| ds\|_2^2$$

and the proof of (2.16) is complete by substituting the bounds on $\|\Delta Y_{t_0}\|_2^2$ from above.

Next, we prove (2.17). Recall that $\mathbb{E}_t[H_s^t] = 0$ for all (t, s), which implies that $\mathbb{E}_t[\Phi_i H_T^t] = \mathbb{E}_t[(\Phi_i - \mathbb{E}_t[\Phi_i]) H_T^t]$. Using the representation $Z_{i,t} := \mathbb{E}_t[\Phi_i H_T^t + \int_t^T f_i(r, Y_{i,r}, Z_{i,r}) H_r^t dr]$, it follows from Minkowski's inequality, the Cauchy-Schwarz inequality (i.e. Corollary 2.11), and Lemma 2.10 that

$$\|\Delta Z_t\|_2 \le \frac{C V_{t,T}(\Delta\Phi)}{\sqrt{T-t}} + C\int_t^T \frac{\|\Delta f_r\|_2}{\sqrt{r-t}} dr + C\int_t^T \frac{\|\Delta Y_r\|_2 + \|\Delta Z_r\|_2}{(T-r)^{(1-\theta_{2,L})/2}\sqrt{r-t}} dr. \quad (2.20)$$

where we define $V_{t,T}(\Delta\Phi)$ by $\mathbb{E}[|\Delta\Phi - \mathbb{E}_t[\Delta\Phi]|^2]^{1/2}$. Defining $\Theta_r := \|\Delta Y_r\|_2 + \|\Delta Z_r\|_2$ and recalling (2.18), it follows that

$$\Theta_t \le C\|\Delta\Phi\|_2 + \frac{C V_{t,T}(\Delta\Phi)}{\sqrt{T-t}} + C\int_t^T \frac{\|\Delta f_r\|_2}{\sqrt{r-t}} dr + C\int_t^T \frac{\Theta_r}{(T-r)^{(1-\theta_{2,L})/2}\sqrt{r-t}} dr. \quad (2.21)$$

Applying Lemma C.3 with $u_t := \Theta_t$ and

$$w_t := C\|\Delta\Phi\|_2 + \frac{CV_{t,T}(\Delta\Phi)}{\sqrt{T-t}} + C\int_t^T \frac{\|\Delta f_r\|_2}{\sqrt{r-t}}dr,$$

it follows that

$$\Theta_r \le Cw_t + C\int_t^T \frac{w_r}{(T-r)^{(1-\theta_{2,L})/2}\sqrt{r-t}}dr + C\int_t^T \frac{\Theta_r}{(T-r)^{(1-\theta_{2,L})/2}}dr$$

whence it follows from Lemma C.4 that

$$\int_t^T \frac{\Theta_r}{(T-r)^{(1-\theta_{2,L})/2}\sqrt{r-t}}dr \le C\int_t^T \frac{w_r}{(T-r)^{(1-\theta_{2,L})/2}\sqrt{r-t}}dr$$

Substituting this into (2.20) and applying Lemma C.2 leads to

$$\begin{aligned}
\|\Delta Z_t\|_2 &\le \frac{CV_{t,T}(\Delta\Phi)}{\sqrt{T-t}} + C\int_t^T \frac{\|\Delta f_r\|_2}{\sqrt{r-t}}dr + C\int_t^T \frac{w_r}{(T-r)^{(1-\theta_{2,L})/2}\sqrt{r-t}}dr \\
&= \frac{CV_{t,T}(\Delta\Phi)}{\sqrt{T-t}} + C\int_t^T \frac{\|\Delta f_r\|_2}{\sqrt{r-t}}dr + C\int_t^T \frac{V_{r,T}(\Delta\Phi)}{(T-r)^{(2-\theta_{2,L})/2}\sqrt{r-t}}dr \\
&\quad + C\int_t^T \frac{\int_r^T \|\Delta f_s\|_2 (s-r)^{-1/2}ds}{(T-r)^{(1-\theta_{2,L})/2}\sqrt{r-t}}dr + C\|\Delta\Phi\|_2(T-t)^{\theta_{2,L}/2} \\
&= \frac{CV_{t,T}(\Delta\Phi)}{\sqrt{T-t}} + C\int_t^T \frac{\|\Delta f_r\|_2}{\sqrt{r-t}}dr + C\int_t^T \frac{V_{r,T}(\Delta\Phi)}{(T-r)^{(2-\theta_{2,L})/2}\sqrt{r-t}}dr \\
&\quad + C\int_t^T \|\Delta f_s\|_2 \{\int_r^s (s-r)^{-1+\theta_{2,L}}(r-t)^{-1/2}dr\}ds + C\|\Delta\Phi\|_2(T-t)^{\theta_{2,L}/2}.
\end{aligned}$$

The proof is completed by observing that $V_{r,T}(\Delta\Phi)$ is non-increasing in r. □

The estimates (2.17) allow us to determine a priori estimates on the conditional second moments of the solution of the BSDE (Y, Z).

Corollary 2.13. *Assume that* $Z_t = \mathbb{E}_t[\Phi(X_T)H_T^t + \int_t^T f(r, X_r, Y_r, Z_r)H_r^t dr]$ *for all* $t \in [0, T)$ *almost surely. Then there is a constant* C *depending only on* L_f, θ_L, C_f, θ_c, $K^\alpha(\Phi)$ *and* T *such that, for all* $t \in [0, T)$ *and* $s \in [0, t]$, *we have*

$$\sup_{s\le t\le T} (\mathbb{E}_s[|Y_t|^2])^{1/2} \le C(1 + (\mathbb{E}_s[|\Phi(X_T) - \mathbb{E}_s[\Phi(X_T)]|^2])^{1/2}),$$

$$(\mathbb{E}_s[|Z_t|^2])^{1/2} \le \frac{C(\mathbb{E}_s[|\Phi(X_T) - \mathbb{E}_s[\Phi(X_T)]|^2])^{1/2}}{\sqrt{T-t}} + \frac{C}{(T-t)^{(1-2\theta_c)/2}} + C(\mathbb{E}_s[|\Phi(X_T)|^2])^{1/2}(T-t)^{\theta_L/2}.$$

In particular, $\|Y_t\|_2 \le C$ *and* $\|Z_t\|_2 \le C(T-s)^{((2\theta_c)\wedge\alpha-1)/2}$ *for all* $t \in [0, T)$, *and*

$$\|f(s, X_s, Y_s, Z_s)\|_2 \le \frac{C}{(T-s)^{1-((2\theta_c)\wedge\alpha+\theta_L)/2}} + \frac{C}{(T-s)^{1-\theta_c}}. \tag{2.22}$$

If $(\mathbf{A_{h\Phi}})$ *is in force, we have additionally that* $|Z_t| \le CK_\Phi(T-s)^{((2\theta_c)\wedge\theta_\Phi-1)/2}$ *for all* $t \in [0, T)$ *almost surely.*

Proof. In what follows, C may change from line to line. As in Proposition 2.12, we only prove the result for $s = 0$; the general case is proved using the conditional version of the Minkowski, Cauchy-Schwarz (Corollary 2.11), and Hölder inequalities in the place of the usual version of these with the regular expectation. Recalling $V_{t,T}(\Phi)$ from $(\mathbf{A_\Phi})$, apply (2.17) from Proposition 2.12 with $(Y_1, Z_1) := (0, 0)$ and $(Y_2, Z_2) := (Y, Z)$ to obtain (for all $t \in [0, T)$)

$$\begin{aligned}\|Z_t\|_2 &\leq C\frac{V_{t,T}(\Phi)}{\sqrt{T-t}} + C\int_t^T \frac{\|f(r, X_r, 0, 0)\|_2}{\sqrt{r-t}}dr + C\|\Phi\|_2(T-t)^{\theta_L/2} \\ &\leq \frac{C}{(T-t)^{(1-\alpha)/2}} + C\int_t^T \frac{dr}{(T-r)^{1-\theta_c}\sqrt{r-t}} + C(T-t)^{\theta_L/2}.\end{aligned}$$

Combining the local Lipschitz continuity and boundedness of f in (1.2) leads to the required bound on the conditional second moments of Z_t. The estimate on the conditional moments of Y_t is obtained similarly starting from (2.16). The remaining bounds are obtained by taking into account (1.2) and the regularity of the terminal condition ($(\mathbf{A_\Phi})$ or $(\mathbf{A_{h\Phi}})$). □

Recall $(Y^{(\varepsilon)}, Z^{(\varepsilon)})$ from Definition 2.7 in Section 2.3, the BSDE with terminal condition Φ and driver $f^{(\varepsilon)}(t, x, y, z) := f(t, x, y, z)\mathbf{1}_{[0,T-\varepsilon)}(t)$. The following corollary of Proposition 2.12 will be used extensively throughout this paper; it provides a stability results between the BSDEs (Y, Z) and $(Y^{(\varepsilon)}, Z^{(\varepsilon)})$ that are controlled by ε.

Corollary 2.14. *Let $\gamma := (\theta_c \wedge \frac{\alpha}{2} + \frac{\theta_L}{2}) \wedge \theta_c$ and assume that $Z = \mathbb{E}_t[\Phi(X_T)H_T^t + \int_t^T f(s, X_s, Y_s, Z_s)H_s^t ds]$ and $Z_t^{(\varepsilon)} = \mathbb{E}_t[\Phi(X_T)H_T^t + \int_t^T f^{(\varepsilon)}(s, X_s, Y_s^{(\varepsilon)}, Z_s^{(\varepsilon)})H_s^t ds]$ for all $t \in [0, T)$ almost surely. Then there is a constant C such that*

$$\sup_{0\leq t\leq T} \|Y_t - Y_t^{(\varepsilon)}\|_2^2 + \int_0^T \|Z_t - Z_t^{(\varepsilon)}\|_2^2 dt \leq C\varepsilon^{2\gamma}, \tag{2.23}$$

$$\|Z_t - Z_t^{(\varepsilon)}\|_2 \leq C\int_{t\vee(T-\varepsilon)}^T \frac{ds}{(T-s)^{1-\gamma}\sqrt{s-t}} \tag{2.24}$$

for all $t \in [0, T)$. In particular, $(Y^{(\varepsilon)}, Z^{(\varepsilon)}) \to (Y, Z)$ as $\varepsilon \to 0$ in $\mathcal{S}^2 \times \mathcal{H}^2$.

Proof. In what follows, C may change from line to line.

It follows from (2.16) in Proposition 2.12 that

$$\sup_{0\leq t\leq T} \|Y_t - Y_t^{(\varepsilon)}\|_2^2 + \int_0^T \|Z_s - Z_s^{(\varepsilon)}\|_2^2 ds \leq C\Big(\int_{T-\varepsilon}^T \|f(s, X_s, Y_s, Z_s)\|_2 ds\Big)^2. \tag{2.25}$$

Substituting (2.22) into (2.25) combined with $\Big(\int_{T-\varepsilon}^T \frac{ds}{(T-s)^{(1-\gamma)}}\Big)^2 \leq C\varepsilon^{2\gamma}$ completes the proof of (2.23). Next, it follows from (2.17) that

$$\|Z_t - Z_t^{(\varepsilon)}\|_2 \leq C\int_{t\vee(T-\varepsilon)}^T \frac{\|f(s, X_s, Y_s, Z_s)\|_2}{\sqrt{s-t}}ds \quad \text{for all } t \in [0, T).$$

Substituting (2.22) above proves (2.24). □

To end this section, we present a mollification procedure that will be used frequently to allow us to extend results under the assumptions $(\mathbf{A_{\partial f}})$ and $(\mathbf{A_{b\Phi}})$ to the same results without these assumptions. The following corollary is a trivial consequence of Proposition 2.12 and the properties of mollifiers.

Corollary 2.15. *Let $M > 0$ be finite, and $M \mapsto R(M) \geq 1$ be increasing w.r.t. M. Define $\Phi_M(x) := -M \vee \Phi(x) \wedge M$ and, recalling the mollifier ϕ of Definition 1.1,*

$$f_M(t, x, y, z) := \int_{\mathbb{R}^d\times\mathbb{R}\times(\mathbb{R}^q)^\top} f(t, x - x', y - y', z - z')\phi_{R(M)}(x', y', z')d(x', y', z'),$$

Let (Y_M, Z_M) be the solution of the BSDE with terminal condition Φ_M and driver $f_M(t, x, y, z)$. Then Φ_M satisfies $(\mathbf{A_{b\Phi}})$, f_M satisfies $(\mathbf{A_{\partial f}})$, and $(Y_M, Z_M) \to (Y, Z)$ as $M \to \infty$ in $\mathcal{S}^2 \times \mathcal{H}^2$.

2.5 Representation theorem

In this section, we prove that BSDEs satisfying the local Lipschitz continuity and local boundedness conditions $(\mathbf{A_f})$ also satisfy the a representation theorem in the spirit of [MZ02, Theorem 3.1]. Following on from Section 2.4, we see that this representation is very valuable, as it gives us additional access to a priori results. We use these a priori results in the sections that follow, so it is essential that we also establish the representation result. Unlike in the proof of [MZ02, Theorem 3.1], we do not prove the representation result on Z directly. The strategy is rather to take the approximative BSDE $(Y^{(\varepsilon)}, Z^{(\varepsilon)})$, for which we already know that $Z^{(\varepsilon)}$ satisfies the representation from [MZ02, Theorem 3.1], then to prove it converges in $\mathcal{H}^2$ to the process that we claim is a version of Z as ε converges to 0 by classical $(\varepsilon, \delta)-$arguments, and to finally conclude using the fact that $Z^{(\varepsilon)}$ also converges to Z in $\mathcal{H}^2$ and because Z is unique.

Theorem 2.16. *Recall $\mathbf{L}_{2,\alpha}$ from $(\mathbf{A_\Phi})$, suppose that $\Phi \in \mathbf{L}_{2,\alpha}$ and $(t,x,y,z) \mapsto f(t,x,y,z)$ satisfies $(\mathbf{A_f})$. Then, there is a predictable version $\mathcal{Z}$ of Z which satisfies*

$$\mathcal{Z}_t = \mathbb{E}_t[\Phi(X_T)H_T^t + \int_t^T f(s, X_s, Y_s, Z_s)H_s^t ds] \quad \textit{for all } t \in [0,T) \quad \mathbb{P}-a.s. \tag{2.26}$$

where H_s^t are the Malliavin weights given in (2.14).

Proof. In the following, C is a constant whose value may change from line to line.

To start with, let assume $(\mathbf{A_{\partial f}})$ and $(\mathbf{A_{b\Phi}})$ be in force. We prove the representation theorem first under these conditions, and then extend to the general result by means of mollification. Recall the BSDEs $(Y^{(\varepsilon)}, Z^{(\varepsilon)})$, (y,z) and $(y^{(\varepsilon)}, z^{(\varepsilon)})$ from Section 2.3, and the decomposition $(Y^{(\varepsilon)}, Z^{(\varepsilon)}) = (y + y^{(\varepsilon)}, z + z^{(\varepsilon)})$. We first prove the that there is a predictable version of $Z^{(\varepsilon)}$ equalling

$$\mathbb{E}_t[\Phi(X_T)H_T^t + \int_t^T f^{(\varepsilon)}(r, X_r, Y_r^{(\varepsilon)}, Z_r^{(\varepsilon)})H_r^t dr] \quad \text{for all } t \in [0,T) \quad \mathbb{P}-a.s. \tag{2.27}$$

In fact, this is an application of [MZ02, Theorem 4.2]; this is not immediately clear, so we make the calculations explicit for the benefit of the reader. Definition 2.7 and Lemma 2.8 give us that $(y^{(\varepsilon)}, z^{(\varepsilon)})$ solves the BSDE with terminal condition 0 and driver

$$F(t,x,y,z) := f^{(\varepsilon)}(t, x, u(t,x) + y, \nabla_x u(t,x)\sigma(t,x) + z)$$

on the time interval $[0, T-\varepsilon]$. Due to the bounds on u and its derivatives given in Lemma 2.8, the Lipschitz constant of $(x,y,z) \mapsto F(t,x,y,z)$ is bounded from above (for all $t \in [0, T-\varepsilon]$) by

$$\frac{L_f}{(T-t)^{(1-\theta_L)/2}}\{1 + \|\nabla_x u(t,\cdot)\|_\infty + \|\nabla_x^2 u(t,\cdot)\|_\infty\} \leq \frac{C}{(T-t)^{(3-\theta_L)/2}} \leq C\varepsilon^{-(3-\theta_L)/2} =: L_F.$$

Using this Lipschitz constant, we also show that $F(t,x,0,0)$ is bounded (for all $(t,x) \in [0, T-\varepsilon] \times \mathbb{R}^d$ by

$$\frac{C_f}{(T-t)^{1-\theta_c}} + L_F\{1 + \|u(t,\cdot)\|_\infty + \|\nabla_x u(t,\cdot)\|_\infty\} \leq C_f\varepsilon^{-(1-\theta_c)} + L_F\varepsilon^{-(1-\theta_L)/2} =: C_F.$$

Therefore, the driver F is uniformly Lipschitz continuous in (x,y,z) and uniformly bounded at $(y,z) = (0,0)$, i.e. it satisfies $(\mathbf{A_f})$ with $\theta_{L,F} \equiv 1$, $\theta_{C,F} \equiv 1$, and constants L_F and C_F (given above). F is also continuous in t. Therefore, [MZ02, Theorem 4.2] applies to the BSDE in the interval $[0, T-\varepsilon]$, i.e. there is a version of $z^{(\varepsilon)}$ equalling

$$\mathbb{E}_t[\int_t^{T-\varepsilon} F(r, X_r, y_r^{(\varepsilon)}, z_r^{(\varepsilon)})H_r^t dr] \ \text{ for all } t \in [0, T-\varepsilon] \ \text{ almost surely.}$$

On the other hand, $z_t^{(\varepsilon)}$ and $F(t,x,y,z)$ are 0 for all $t \in (T-\varepsilon, T]$ almost surely, so the representation holds trivially in the interval $(T-\varepsilon, T]$, whence it follows that there is a version of $z^{(\varepsilon)}$ equalling

$$\mathbb{E}_t[\int_t^{T-\varepsilon} F(r, X_r, y_r^{(\varepsilon)}, z_r^{(\varepsilon)}) H_r^t dr] \text{ for all } t \in [0,T] \text{ almost surely.}$$

Now, it is well known – see for example [GM10, Page 1116], where our H_T^t is given by $H_{t,T}^{(1)}\sigma(t, X_t)$ in their notation – that there is predictable version of $(z_t)_{t\in[0,T)}$ equalling

$$\mathbb{E}_t[\Phi(X_T)H_T^t] \text{ for all } t \in [0,T) \text{ almost surely,}$$

and this implies the version of $Z^{(\varepsilon)}$ given by (2.27) thanks to the the decomposition $(Y^{(\varepsilon)}, Z^{(\varepsilon)}) = (y + y^{(\varepsilon)}, z + z^{(\varepsilon)})$.

Define by $\mathcal{Z}$ the predictable projection [JS03, Theorem 2.28] of the process $(\mathcal{X}_t := \Phi(X_T)H_T^t + \int_t^T f(r, X_r, Y_r, Z_r) H_r^t dr)_{t\in[0,T)}$. In what follows, we show that $\|Z_t^{(\varepsilon)} - \mathcal{Z}_t\|_2 \to 0$ as $\varepsilon \to 0$ for almost all $t \in [0,T)$. This implies, by the dominated convergence theorem, that $Z^{(\varepsilon)} \to \mathcal{Z}$ in $\mathcal{H}^2$. Since $Z^{(\varepsilon)} \to Z$ in $\mathcal{H}^2$ was determined in Corollary 2.14, this implies that $Z_t = \mathcal{Z}_t$ $m \times \mathbb{P} - a.e.$, which completes the proof under the assumptions $(\mathbf{A_{\partial f}})$ and $(\mathbf{A_{b\Phi}})$.

We first need some intermediate upper bounds. Analogously to Corollary 2.13, we have that

$$\|f(r, X_r, Y_r^{(\varepsilon)}, Z_r^{(\varepsilon)})\|_2 \le \frac{C}{(T-r)^{1-\gamma}} \quad \text{for all } r \in [0,T). \tag{2.28}$$

Fix $t \in [0,T)$ and $\eta > 0$. Using the representation formula (2.27), it follows from Minkowski's inequality, the conditional Cauchy-Schwarz inequality (Corollary 2.11), and Lemma 2.10 that

$$\begin{aligned}
\|Z_t^{(\varepsilon)} - \mathcal{Z}_t\|_2 &= \|\mathbb{E}_t[\int_t^T \big(f^{(\varepsilon)}(r, X_r, Y_r^{(\varepsilon)}, Z_r^{(\varepsilon)}) - f(r, X_r, Y_r, Z_r)\big) H_r^t dr]\|_2 \\
&\le \|\mathbb{E}_t[\int_t^T \big(f^{(\varepsilon)}(r, X_r, Y_r^{(\varepsilon)}, Z_r^{(\varepsilon)}) - f(r, X_r, Y_r^{(\varepsilon)}, Z_r^{(\varepsilon)})\big) H_r^t dr]\|_2 \\
&\quad + C_M^{1/2} \int_t^T \frac{\|f(r, X_r, Y_r^{(\varepsilon)}, Z_r^{(\varepsilon)}) - f(r, X_r, Y_r, Z_r)\|_2}{\sqrt{r-t}} dr.
\end{aligned} \tag{2.29}$$

Taking $\varepsilon < (T-t)/2$ and using (2.28), it follows that

$$\begin{aligned}
\|\mathbb{E}_t[\int_t^T \big(f^{(\varepsilon)}(r, X_r, Y_r^{(\varepsilon)}, Z_r^{(\varepsilon)}) - f(r, X_r Y_r^{(\varepsilon)}, Z_r^{(\varepsilon)})\big) H_r^t dr]\|_2 &\le C_M^{1/2} \int_{T-\varepsilon}^T \frac{\|f(r, X_r, Y_r^{(\varepsilon)}, Z_r^{(\varepsilon)})\|_2}{\sqrt{r-t}} dr \\
\le \frac{C_M^{1/2}}{\sqrt{T-t-\varepsilon}} \int_{T-\varepsilon}^T \|f(r, X_r, Y_r^{(\varepsilon)}, Z_r^{(\varepsilon)})\|_2 dr \le \frac{C}{\sqrt{T-t}} \int_{T-\varepsilon}^T \frac{dr}{(T-r)^{1-\gamma}} &= \frac{C\varepsilon^\gamma}{\sqrt{T-t}}.
\end{aligned}$$

Taking $\varepsilon < \eta^{1/\gamma}(T-t)^{1/(2\gamma)}/C$, where C is the last constant in the inequality above, is sufficient to bound the above term by η. On the other hand, letting $\delta < (T-t)/2$,

$$\begin{aligned}
&\int_t^T \frac{\|f(r, X_r, Y_r^{(\varepsilon)}, Z_r^{(\varepsilon)}) - f(r, X_r, Y_r, Z_r)\|_2}{\sqrt{r-t}} dr \\
&\le C_M^{1/2} \frac{\int_{t+\delta}^T \|f(r, X_r, Y_r^{(\varepsilon)}, Z_r^{(\varepsilon)}) - f(r, X_r, Y_r, Z_r)\|_2 dr}{\sqrt{\delta}} \\
&\quad + C_M^{1/2} \int_t^{t+\delta} \frac{\|f(r, X_r, Y_r^{(\varepsilon)}, Z_r^{(\varepsilon)}) - f(r, X_r, Y_r, Z_r)\|_2}{\sqrt{r-t}} dr
\end{aligned} \tag{2.30}$$

To bound the first integral term on the right hand side above, we apply Hölder's inequality and the Lipschitz continuity of $f(t,\cdot)$ to obtain

$$
\begin{aligned}
&C_M^{1/2}\int_{t+\delta}^{T}\|f(r,X_r,Y_r^{(\varepsilon)},Z_r^{(\varepsilon)})-f(r,X_r,Y_r,Z_r)\|_2 dr\\
&\qquad \le C_M^{1/2}L_f\Big(\int_0^T\frac{dr}{(T-r)^{1-\theta_L}}\Big)^{1/2}\Big(\sup_{0\le s\le T}\|Y_s-Y_s^{(\varepsilon)}\|_2^2+\int_0^T\|Z_r-Z_r^{(\varepsilon)}\|_2^2 dr\Big)^{1/2}
\end{aligned}
$$

Using that $(Y^{(\varepsilon)},Z^{(\varepsilon)})\to(Y,Z)$ in $\mathcal{S}\times\mathcal{H}^2$ as $\varepsilon\to 0$ (Corollary 2.14), set ε sufficiently small so that the above is bounded above by $\sqrt{\delta}\eta$. To bound the second integral term on the right hand side of (2.30), we use (2.22) and (2.28) combined with the triangle inequality to show that

$$
\begin{aligned}
&C_M^{1/2}\int_t^{t+\delta}\frac{\|f(r,X_r,Y_r^{(\varepsilon)},Z_r^{(\varepsilon)})-f(r,X_r,Y_r,Z_r)\|_2}{\sqrt{r-t}}dr\\
&\qquad\le\frac{C}{(T-t-\delta)^{1-\gamma}}\int_t^{t+\delta}\frac{dr}{\sqrt{r-t}}\le\frac{C\sqrt{\delta}}{(T-t)^{1-\gamma}}
\end{aligned}
$$

and set δ sufficiently small so that the above is bounded above by η. Therefore, we have shown that for almost every $t\in[0,T)$ and every $\eta>0$, there is a sufficiently small ε such that $\|Z_t^{(\varepsilon)}-\mathcal{Z}_t\|_2<3\eta$. In other words, $\mathbb{E}[|Z_t^{(\varepsilon)}-\mathcal{Z}_t|^2]\to 0$ as $\varepsilon\to 0$ for every t, as required.

To prove the result without $(\mathbf{A_{\partial f}})$ and $(\mathbf{A_{b\Phi}})$, recall the mollified BSDE (Y_M,Z_M) from Corollary 2.15. Since Φ_M satisfies $(\mathbf{A_{b\Phi}})$ and f_M satisfies $(\mathbf{A_{\partial f}})$, there is a predictable version $\mathcal{Z}_M$ of Z_M satisfying $\mathcal{Z}_{M,t}=\mathbb{E}_t[\Phi_M(X_T)H_T^t+\int_t^T f_M(r,X_r,Y_{M,r},Z_{M,r})H_r^t dr]$ for all $t\in[0,T)$ almost surely. Thanks to the point-wise convergence of f_M to f and Φ_M to Φ, and the convergence of (Y_M,Z_M) to (Y,Z) in $\mathcal{S}^2\times\mathcal{H}^2$ from Corollary 2.15, we can use analogous limit arguments as above to complete the proof. □

3 Convergence rate of the Euler scheme for BSDEs

Throughout this section, the assumption $(\mathbf{A_{f_t}})$ is in force. Let us recall now the Euler scheme for BSDEs:

$$
\begin{aligned}
Y_N^{(N)}&:=\Phi(X_T),\quad Z_i^{(N)}:=\frac{1}{t_{i+1}-t_i}\mathbb{E}_i[Y_{i+1}^{(N)}(W_{t_{i+1}}-W_{t_i})^\top],\\
Y_i^{(N)}&:=\mathbb{E}_i[Y_{i+1}^{(N)}+f(t_i,X_{t_i},Y_{i+1}^{(N)},Z_i^{(N)})(t_{i+1}-t_i)].
\end{aligned}
$$

We determine error estimates on the error of the Euler scheme, which is given by

$$
\mathcal{E}(N):=\max_{0\le i<N}\mathbb{E}[|Y_{t_i}-Y_i^{(N)}|^2]+\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\mathbb{E}[|Z_t-Z_i^{(N)}|^2]dt.
$$

The following proposition serves as the starting point of our analysis; it allows us to estimate the error $\mathcal{E}(N)$ using estimates for the so called $\mathbf{L}_2$-regularity, which we will do subsequently.

Proposition 3.1. *Let $\beta\le\theta_L$. For the Euler scheme for BSDEs defined on the time-grids $\{\pi_N^{(\beta)}:N\ge 1\}$, there is a constant C depending only on L_f, L_X, θ_L, θ_X, β, and T, but not on N, such that, for all $N\ge 1$,*

$$
\mathcal{E}(N)\le CN^{-1}+C\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\|Z_t-\tilde{Z}_{t_i}\|_2^2 dt
$$

where $\tilde{Z}_{t_i}:=\frac{1}{\Delta_i}\mathbb{E}_i\big[\int_{t_i}^{t_{i+1}}Z_t dt\big]$.

The proof is analogous to the proof of [GL06, Theorem 1], one must only use the result $\Delta_k/(T-t_k)^{1-\theta_L} \leq T^{\theta_L}(\beta N)^{-1}$ for $\beta \leq \theta_L$ (see Lemma B.1) in order to compensate for the local Lipschtz constant of the driver.

The sum $\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \|Z_t - \tilde{Z}_{t_i}\|_2^2 dt$ is called the $\mathbf{L}_2$-regularity; it's study was initiated by [Zha04]. Since $(\tilde{Z}_{t_i} := \frac{1}{\Delta_i}\mathbb{E}_i\big[\int_{t_i}^{t_{i+1}} Z_t dt\big])_i$ is the projection of Z onto the space of adapted discrete processes with nodes on π under the scalar product $(u,v) = \mathbb{E}\int_0^T (u_s \cdot v_s) ds$, it follows that

$$\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \|Z_t - \tilde{Z}_{t_i}\|_2^2 dt \leq \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \|Z_t - Z_{t_i}\|_2^2 dt. \tag{3.1}$$

To bound $\mathcal{E}(N)$, it follows from Proposition 3.1 that it is sufficient to bound the term on the right-hand side of (3.1). However, as in the proof of the Representation Theorem in Section 2.5, it is not possible to do so directly for the BSDE (Y, Z), so we use an approximation procedure via the BSDE $(Y^{(\varepsilon)}, Z^{(\varepsilon)})$, which we recall from Definition 2.7 in Section 2.3.

Throughout the remainder of this section, we work with the version of Z and $Z^{(\varepsilon)}$ given by Theorem 2.16, i.e

$$Z_t = \mathbb{E}_t[\Phi(X_T)H_T^t + \int_t^T f(s, X_s, Y_s, Z_s)H_s^t ds] \quad \text{for all } t \in [0,T) \quad \mathbb{P}-a.s.,$$

$$Z_t^{(\varepsilon)} = \mathbb{E}_t[\Phi(X_T)H_T^t + \int_t^T f^{(\varepsilon)}(r, X_r, Y_r^{(\varepsilon)}, Z_r^{(\varepsilon)})H_r^t dr] \quad \text{for all } t \in [0,T) \quad \mathbb{P}-a.s.$$

This version empowers us with the additional a priori estimates estimates developed in Section 2.4; we use these estimates frequently in the analysis of this section.

The following lemma decomposes the $\mathbf{L}_2$-regularity of Z – the left hand side of equation (3.1) – into the $\mathbf{L}_2$-regularity of $Z^{(\varepsilon)}$ and a small correction term controlled by ε.

Lemma 3.2. *Let $\beta \in (0,1]$. Then there is a constant C depending only on L_f, C_M, θ_L, θ_c, β, C_f, $K^\alpha(\Phi)$, and T, such that for all $N \geq 1$*

$$\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \|Z_s - \tilde{Z}_{t_i}\|_2^2 ds \leq C \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \|Z_s^{(\varepsilon)} - \tilde{Z}_{t_i}^{(\varepsilon)}\|_2^2 ds + C\varepsilon^{2\gamma}$$

where $\tilde{Z}_{t_i} := \frac{1}{\Delta_i}\mathbb{E}_i\big[\int_{t_i}^{t_{i+1}} Z_t dt\big]$, $\tilde{Z}_{t_i}^{(\varepsilon)} := \frac{1}{\Delta_i}\mathbb{E}_i\big[\int_{t_i}^{t_{i+1}} Z_t^{(\varepsilon)} dt\big]$, and $\gamma := (\theta_c \wedge \frac{\alpha}{2} + \frac{\theta_L}{2}) \wedge \theta_c$.

Proof. In what follows, C may change in value from line to line. Using the Cauchy inequality and the orthogonality of the projections, $\frac{1}{2}\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} \|Z_s - \tilde{Z}_{t_i}\|_2^2 ds \leq \int_0^T \|Z_s - Z_s^{(\varepsilon)}\|_2^2 ds + \sum_{i=0}^{N-1} \|\tilde{Z}_{t_i} - \tilde{Z}_{t_i}^{(\varepsilon)}\|_2^2 \Delta_i + \sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} \|Z_s^{(\varepsilon)} - \tilde{Z}_{t_i}^{(\varepsilon)}\|_2^2 ds$. Recall from Corollary 2.14 that $\int_0^T \|Z_s - Z_s^{(\varepsilon)}\|_2^2 ds \leq C\varepsilon^{2\gamma}$. Moreover, using Jensen's inequality,

$$\sum_{i=0}^{N-1} \|\tilde{Z}_{t_i} - \tilde{Z}_{t_i}^{(\varepsilon)}\|_2^2 \Delta_i = \sum_{i=0}^{N-1} \|\frac{1}{\Delta_i}\mathbb{E}_i\big[\int_{t_i}^{t_{i+1}} (Z_t - Z_t^{(\varepsilon)}) dt\|_2^2 \Delta_i \leq \int_0^T \|Z_s - Z_s^{(\varepsilon)}\|_2^2 ds \leq C\varepsilon^{2\gamma}$$

and this completes the proof. □

We now come to our first and most general estimate on the $\mathcal{E}(N)$. Later, in Theorem 4.5, we augment this result with stronger assumptions.

Theorem 3.3. *Let $0 < \beta < (2\gamma) \wedge \alpha$ and $\gamma := (\theta_c \wedge \frac{\alpha}{2} + \frac{\theta_L}{2}) \wedge \theta_c$. There is a constant C depending only on L_f, C_M, θ_L, θ_c, β, C_f, $K^\alpha(\Phi)$, and T, but not on N, such that for all $N \geq 1$,*

$$\mathcal{E}(N) \leq CN^{-1}\mathbf{1}_{[1,2]}(\alpha + \theta_L) + CN^{-2\gamma}\mathbf{1}_{(0,1)}(\alpha + \theta_L)$$

Proof. In what follows, C may change in value from line to line. From Proposition 3.1, it is sufficient to bound $\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\|Z_t-\tilde{Z}_{t_i}\|_2^2dt$. To start with, assume $(\mathbf{A_{\partial f}})$ and $(\mathbf{A_{b\Phi}})$. Recall the BSDEs $(y^{(\varepsilon)},z^{(\varepsilon)})$ from Definition 2.7 and $(U^{(\varepsilon)},V^{(\varepsilon)})$ from (2.9) in Section 2.3. In the proof of [GM10, Theorem 3.1], the authors show that for any i and $s\in[t_i,t_{i+1})$,

$$\|z_s^{(\varepsilon)}-z_{t_i}^{(\varepsilon)}\|_2\le C\int_{t_i}^{s}\|a_r^{(\varepsilon)}\|_2dr+C\int_{t_i}^{s}\|V_r^{(\varepsilon)}\|_2dr+C\Delta_i^{1/2}. \tag{3.2}$$

Using $(\int_0^T\|a_r^{(\varepsilon)}\|_2dr)^2+\int_0^T\|V_r^{(\varepsilon)}\|_2^2dr\le C\varepsilon^{-1+(\theta_L+\alpha)\wedge 1}$ from (2.10) in Lemma 2.9, and (3.1), it follows from Jensen's inequality that

$$\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\|z_s^{(\varepsilon)}-\tilde{z}_{t_i}^{(\varepsilon)}\|_2^2ds\le\frac{C}{N}+\frac{C\max_{0\le i\le N-1}\Delta_i}{\varepsilon^{1-(\theta_L+\alpha)\wedge 1}}\le CN^{-1}(1+\varepsilon^{(\theta_L+\alpha)\wedge 1-1})$$

where $\max_i\Delta_i\le CN^{-1}$ follows from (B.1) in Lemma B.1. Combining this estimate with $\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\|z_s-z_{t_i}\|_2^2ds\le CN^{-1}$, shown in [GM10, Theorem 1.3], $Z^{(\varepsilon)}=z+z^{(\varepsilon)}$, and the results of Lemma 3.2, (3.1) it follows that

$$\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\|Z_s-\tilde{Z}_{t_i}\|_2^2ds\le CN^{-2\gamma/\beta}+C\varepsilon^{2\gamma}+CN^{-1}(1+\varepsilon^{(\theta_L+\alpha)\wedge 1-1}). \tag{3.3}$$

To complete the proof under $(\mathbf{A_{\partial f}})$ and $(\mathbf{A_{b\Phi}})$, let $\varepsilon:=N^{-\delta}$ in the estimate (3.3), take $\delta:=1/(2\gamma)$ if $\alpha+\theta_L\ge 1$ and $\delta:=1$ otherwise, and notice that $2\gamma\le\alpha+\theta_L$.

In order to prove the general result, recall the BSDE (Y_M,Z_M) from Corollary 2.15; its terminal condition satisfies $(\mathbf{A_{b\Phi}})$ and its driver satisfies $(\mathbf{A_{\partial f}})$. Moreover, [GM10, Lemma 3.1] proves $K^\alpha(\Phi_M)\le K^\alpha(\Phi)$. Therefore, working with the version of Z_M given by the representation formula $Z_{M,t}=\mathbb{E}_t[\Phi_M(X_T)H_T^t+\int_t^T f_M(s,X_s,Y_{M,s},Z_{M,s})H_s^tds]$ from Theorem 2.16, it follows from the triangle inequality and the results obtained above that

$$\mathcal{E}(N)\le 2\int_0^T\|Z_s-Z_{M,s}\|_2^2ds+2\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\|Z_{M,s}-\tilde{Z}_{M,t_i}\|_2^2ds$$

and letting $M\to\infty$ with Corollary 2.15 yields the result. $\square$

4 A priori estimates under $(\mathbf{A_{b\Phi}})$ and $(\mathbf{A_{h\Phi}})$

At the end this section, we give a complementary result to Theorem 3.3 under stronger the conditions on the terminal condition $(\mathbf{A_{b\Phi}})$ and $(\mathbf{A_{h\Phi}})$, i.e. where the function Φ is bounded (and/)or Hölder continuous, respectively. This is achieved using the an additional a priori estimates on $\|V_t^{(\varepsilon)}\|_2$, given in Proposition 4.2 below. Moreover, these a priori estimates will be critical in Section 5, where one requires more structure than in Section 3. The result is proved, roughly speaking, by using a functional representation of the intermediate process $z^{(\varepsilon)}$ and show Lipschitz continuity of the said functional representation. This adds an additional layer of interest under $(\mathbf{A_{h\Phi}})$ for the parameters $\theta_\Phi+\theta_L\ge 1$, where we can demonstrate that limit of the process $z_s^{(\varepsilon)}$ in $\mathcal{H}^2$, i.e. the process $Z_s-\nabla_x u(s,X_s)\sigma(s,X_s)$, has a functional representation and that function is Lipschitz continuous; see Corollary 4.3. Regularity results are important for numerical schemes as they allow one to build algorithms with lower numerical complexity – see for example [GT13a, Section 3.5] – and this regularity result has such implications for the proxy scheme described in the introduction of this paper.

First, we state the result that $x \mapsto \sigma^{-1}(t,x)$ is uniformly Lipschitz continuous, and $t \mapsto \sigma^{-1}(t,x)$ is uniformly 1/2-Hölder continuous. This elementary result will also be useful in Section 5 below. The proof is to be found in Appendix D.

Lemma 4.1. *The right inverse matrix $\sigma(t,\cdot)^{-1}$ is Lipschitz continuous uniformly in t and $\sigma^{-1}(\cdot,x)$ is 1/2-Hölder continuous uniformly in x. Its Lipschitz (resp. Hölder) constant depends $\|\sigma\|_\infty$, $\|\nabla_x\sigma\|_\infty$ and $\bar\beta$ only, but not on (t,x). Moreover, $\|\sigma^{-1}\|_\infty \le \|\sigma\|_\infty/\bar\beta$.*

We now state the main result of this section, the a priori estimates on the process $V^{(\varepsilon)}$.

Proposition 4.2. *Suppose that $(\mathbf{A_{\partial f}})$ is in force and $\Phi(x)$ is not zero everywhere in $\mathbb{R}^d$. If $(\mathbf{A_{b\Phi}})$ is in force, there exists version of $V^{(\varepsilon)}$ and a finite constant C depending only on L_f, the bounds on b and σ and their partial derivatives, $\bar\beta$, C_M, θ_L, θ_c, C_f, and T such that for any $\varepsilon \in (0,T]$ and every $t \in [0,T)$, $\|V_t^{(\varepsilon)}\|_2 \le C\phi(t,\varepsilon,\theta_L)$, where*

$$\phi(t,\varepsilon,\theta_L) := \|\Phi\|_\infty \int_t^{T-\varepsilon} \frac{dr}{(T-r)^{(3-\theta_L)/2}\sqrt{r-t}}. \tag{4.1}$$

If $(\mathbf{A_{h\Phi}})$ is in force, there exists a version of $V^{(\varepsilon)}$, such that for any $\varepsilon \in (0,T]$ and every $t \in [0,T)$, $\|V_t^{(\varepsilon)}\|_2 \le C\phi(t,\varepsilon,\theta_L,\theta_\Phi)$, where

$$\phi(t,\varepsilon,\theta_L,\theta_\Phi) := K_\Phi \int_t^{T-\varepsilon} \frac{dr}{(T-r)^{(3-\theta_L-\theta_\Phi)/2}\sqrt{r-t}}. \tag{4.2}$$

Remark. The integrals in (4.1,4.2) exist and are bounded by $C\varepsilon^{-(1-\theta_L)/2}(T-t)^{(\alpha-1)/2}$.

Proof. In what follows, C may change from line to line.
Step 1. Functional and BSDE setup. For all $(t,x) \in [0,T) \times \mathbb{R}^d$, consider the FBSDE

$$y_s^{(\varepsilon,t,x)} = \int_s^T F(r, X_r^{(t,x)}, y_r^{(\varepsilon,t,x)}, z_r^{(\varepsilon,t,x)})dr - \int_s^T z_r^{(\varepsilon,t,x)} dW_r, \qquad s \in [t,T), \tag{4.3}$$

where $F(t,x,y,z) = f^{(\varepsilon)}(t,x,u(t,x)+y,(\nabla_x u(t,x)\sigma(t,x))^\top + z)$ and $X^{(t,x)}$ is the solution of the SDE (2.2). Note that the BSDE $(y^{(\varepsilon)}, z^{(\varepsilon)})$ from Section 2.3 is equal to $(y^{(\varepsilon,0,x_0)}, z^{(\varepsilon,0,x_0)})$ because, thanks to Lemma 2.8, (y,z) is equal to $\big(u(\cdot,X_\cdot),\nabla_x u(\cdot,X_\cdot)\sigma(\cdot,X_\cdot)\big)$ and X is equal to $X^{(0,x_0)}$. Since $f^{(\varepsilon)}(t,\cdot)$ is Lipschitz continuous for all $t \in [0,T]$, $F(t,\cdot)$ is also Lipschitz continuous, with Lipschitz constant $C\mathbf{1}_{[0,T-\varepsilon)}(t)\varepsilon^{(\theta_L-3)/2}$, for all $t \in [0,T)$; see the first paragraph of the proof of Theorem 2.16 for detailed computations. Now, letting

$$H_r^{(t,x,s)} := \frac{\mathbf{1}_{(t,T]}(s)}{r-s}\Big(\int_s^r \sigma^{-1}(r,X_r^{(t,x)})D_sX_r^{(t,x)})^\top dW_r\Big)^\top$$

where $D_sX^{(t,x)}$ is the Malliavin derivative of $X_s^{(t,x)}$, it follows from [MZ02, Theorem 4.2], because the terminal condition of the BSDE satisfied by $(y^{(\varepsilon,t,x)}, z^{(\varepsilon,t,x)})$ is zero, that $z_r^{(\varepsilon,t,x)}$ is equal to $z^{(\varepsilon)}(r,X_r^{(t,x)})$ $m \times \mathbb{P}$-almost everywhere, where $z^{(\varepsilon)} : [0,T) \times \mathbb{R}^d \to (\mathbb{R}^q)^\top$ is a continuous, deterministic function given by

$$z^{(\varepsilon)}(t,x) := \mathbb{E}\Big[\int_t^T F(r, X_r^{(t,x)}, y_r^{(\varepsilon,t,x)}, z_r^{(\varepsilon,t,x)})H_r^{(t,x,t)}dr\Big]; \tag{4.4}$$

we work with this version of $z^{(\varepsilon,t,x)}$ from hereon. Additionally, we show in Step 3 below that the process

$$((\nabla X_s^{(t,x)})^\top \nabla_x z^{(\varepsilon)}(s, X_s^{(t,x)}))_{0\le s\le T}$$

(the derivative here is in the weak sense) is a version of the process $(\nabla z_s^{(\varepsilon,t,x)})_{0\le s\le T}$, which is a part of the solution $(\nabla y^{(\varepsilon,t,x)}, \nabla z^{(\varepsilon,t,x)})$ of the BSDE

$$\nabla y_\tau^{(\varepsilon,t,x)} = \int_\tau^T \nabla_x f^{(\varepsilon)}(\Theta_r)\nabla X_r^{(t,x)} + \nabla_y f^{(\varepsilon)}(\Theta_r)\{\nabla_x u(r, X_r^{(t,x)})\nabla X_r^{(t,x)} + \nabla y_r^{(\varepsilon,t,x)}\}dr$$
$$+ \int_\tau^T \nabla_z f^{(\varepsilon)}(\Theta_r) U(r, X_r^{(t,x)})^\top \nabla X_r^{(t,x)} + \sum_{j=1}^q \nabla_z f_j^{(\varepsilon)}(\Theta_r)(\nabla z_{j,r}^{(\varepsilon,t,x)})^\top dr - \sum_{j=1}^q \int_\tau^T (\nabla z_{j,r}^{(\varepsilon,t,x)})^\top dW_r, \tag{4.5}$$

where $\Theta_r = (r, X_r^{(t,x)}, Y_r^{(\varepsilon,t,x)}, Z_r^{(\varepsilon,t,x)})$; the function $U(t,x)$ is defined

$$U(t,x) := \nabla_x^2 u(t,x)\sigma(t,x) + \sum_{j=1}^d (\nabla_x u)_j(t,x)\nabla_x \sigma_j^\top(t,x)$$

for the function u defined in Lemma 2.8. Note that the BSDE (4.5) is a generalization to the BSDE (2.11) – solved by $(\nabla y^{(\varepsilon)}, \nabla z^{(\varepsilon)})$ – which we recall for convenience:

$$\nabla y_t^{(\varepsilon)} = \int_t^T \nabla_x f^{(\varepsilon)}(\Theta_r)\nabla X_r + \nabla_y f^{(\varepsilon)}(\Theta_r)(\nabla_x u(r, X_r)\nabla X_r + \nabla y_r^{(\varepsilon)})dr$$
$$+ \int_t^T \nabla_z f^{(\varepsilon)}(\Theta_r)U(r, X_r)^\top \nabla X_r + \sum_{j=1}^q \nabla_z f_j^{(\varepsilon)}(\Theta_r)(\nabla z_{j,r}^{(\varepsilon)})^\top dr - \sum_{j=1}^q \int_t^T (\nabla z_{j,r}^{(\varepsilon)})^\top dW_r;$$

indeed, in (2.11), set $t \equiv 0$ and $x \equiv x_0$.

Step 2. Proof assuming $z^{(\varepsilon)}(t,\cdot)$ is Lipschitz continuous with $|\nabla_x z^{(\varepsilon)}(t,\cdot)| \le C\phi(t,\cdot)$ and $\big(\nabla X_s^{(t,x)}\big)^\top \nabla_x z^{(\varepsilon)}(s, X_s^{(t,x)})$ is a version of $\nabla z_s^{(\varepsilon,t,x)}$. The hypothesis $|\nabla_x z^{(\varepsilon)}(t,\cdot)| \le C\phi(t,\cdot)$ implies that

$$\|\nabla z_t^{(\varepsilon)}\|_2 = \|(\nabla X_t)^\top \nabla_x z^{(\varepsilon)}(t, X_t)\|_2 \le \|\nabla X_t\|_2 \|\nabla_x z^{(\varepsilon)}(t,\cdot)\|_\infty \le C\phi(t,\cdot) \quad \text{for all } s.$$

Now, using Lemma 2.9,

$$\|\sup_{s\le r<T} U_r^{(\varepsilon)}\|_2 \le C\int_s^{T-\varepsilon} \|a_r^{(\varepsilon)}\|_2 dr \le C\int_s^{T-\varepsilon} \frac{dr}{(T-r)^{(3-\alpha)/2}} \le C\phi(s,\cdot),$$

and $(V_{j,t}^{(\varepsilon)})^\top = (\nabla z_{j,t}^{(\varepsilon)})^\top \sigma^{-1}(t, X_t) - U_t^{(\varepsilon)}\nabla_x \sigma_j(t, X_t)$, therefore we conclude that $\|V_t^{(\varepsilon)}\|_2 \le C\phi(s,\cdot)$ as required.

Step 3. Proving that $\big(\nabla X_s^{(t,x)}\big)^\top \nabla_x z^{(\varepsilon)}(s, X_s^{(t,x)})$ is a version of $\nabla z_s^{(\varepsilon,t,x)}$. We make use of Malliavin calculus – see Section 2.1. By taking the Malliavin derivative on both the BSDE solution $(y^{(\varepsilon)}, z^{(\varepsilon)})$ and on the functional representation $z^{(\varepsilon)}(s, X_s^{(t,x)})$, we obtain an intermediate version that is equal for both.

▶ **BSDE arguments.** There is a version (see [GM10, Lemma 2.2] for the proof) of the processes $(D_s y_\tau^{(\varepsilon,t,x)}, D_s z_\tau^{(\varepsilon,t,x)})_{s\le\tau\le T}$, the Malliavin derivatives of the processes $(y^{(\varepsilon)}, z^{(\varepsilon)})$, solving the BSDE

$$D_s y_\tau^{(\varepsilon,t,x)} = \int_\tau^T \nabla_x f^{(\varepsilon)}(\Theta_r) D_s X_r^{(t,x)} + \nabla_y f^{(\varepsilon)}(\Theta_r)(\nabla_x u(r, X_r) D_s X_r^{(t,x)} + D_s y_r^{(\varepsilon,t,x)})dr$$
$$+ \int_t^T \nabla_z f^{(\varepsilon)}(\Theta_r) U(r, X_r^{(t,x)})^\top D_s X_r^{(t,x)} + \sum_{j=1}^q \nabla_z f_j^{(\varepsilon)}(\Theta_r)(D_s z_{j,r}^{(\varepsilon,t,x)})^\top dr$$
$$- \sum_{j=1}^q \int_t^T (D_s z_{j,r}^{(\varepsilon,t,x)})^\top dW_r. \tag{4.6}$$

We multiply (4.6) on the right by $\sigma^{-1}(s, X_s^{(t,x)})\nabla X_s^{(t,x)}$ and apply Lemma 2.5 to obtain

$$
\begin{aligned}
D_s y_\tau^{(\varepsilon,t,x)}\sigma^{-1}(s, X_s^{(t,x)})\nabla X_s^{(t,x)} &= \int_\tau^T \nabla_x f^{(\varepsilon)}(\Theta_r)\nabla X_r^{(t,x)} + \nabla_y f^{(\varepsilon)}(\Theta_r)(\nabla_x u(r, X_r^{(t,x)})\nabla X_r^{(t,x)})dr \\
&+ \int_\tau^T \nabla_y f^{(\varepsilon)}(\Theta_r) D_s y_r^{(\varepsilon,t,x)}\sigma^{-1}(s, X_s^{(t,x)})\nabla X_s^{(t,x)} dr \\
&+ \int_\tau^T \nabla_z f^{(\varepsilon)}(\Theta_r) U(r, X_r^{(t,x)})^\top \nabla X_r^{(t,x)} dr \\
&+ \sum_{j=1}^q \int_\tau^T \nabla_z f_j^{(\varepsilon)}(\Theta_r)(\big(\sigma^{-1}(s, X_s^{(t,x)})\nabla X_s^{(t,x)}\big)^\top D_s z_{j,r}^{(\varepsilon,t,x)})^\top dr \\
&- \sum_{j=1}^q \int_\tau^T (\big(\sigma^{-1}(s, X_s^{(t,x)})\nabla X_s^{(t,x)}\big)^\top D_s z_{j,r}^{(\varepsilon,t,x)})^\top dW_r; \qquad (4.7)
\end{aligned}
$$

comparing the BSDE (4.7) to to (4.5) term by term, it is clear that

$$
(D_s y_\tau^{(\varepsilon,t,x)}\sigma^{-1}(s, X_s^{(t,x)})\nabla X_s^{(t,x)}, \big(\sigma^{-1}(s, X_s^{(t,x)})\nabla X_s^{(t,x)}\big)^\top D_s z_\tau^{(\varepsilon,t,x)})_{s\le\tau\le T},
$$

a version of the solution to (4.7), is a version of $(\nabla y_\tau^{(\varepsilon,t,x)}, \nabla z_\tau^{(\varepsilon,t,x)})_{s\le\tau\le T}$, the solution to (4.5), for all $s \in [0, T]$.

▶ **Functional arguments.** We start by assuming that $z^{(\varepsilon)}(t, \cdot)$ is smooth (or by taking a mollification). The chain-rule of Malliavin calculus – Lemma 2.1 – yields $D_s z^{(\varepsilon)}(\tau, X_\tau^{(t,x)})$ equals $(D_s X_\tau^{(t,x)})^\top \nabla_x z^{(\varepsilon)}(\tau, X_\tau^{(t,x)})$, and, applying Lemma 2.5, $\big(\sigma^{-1}(s, X_s^{(t,x)})\nabla X_s^{(t,x)}\big)^\top D_s z^{(\varepsilon)}(\tau, X_\tau^{(t,x)})$ is equal to $(\nabla X_\tau^{(t,x)})^\top \nabla_x z^{(\varepsilon)}(\tau, X_\tau^{(t,x)})$. The result follows for $z^{(\varepsilon)}(\tau, \cdot)$ only Lipschitz continuous by standard limiting arguments. Since $(z^{(\varepsilon)}(\tau, X_\tau^{(t,x)}))_{0\le\tau\le T}$ is a version of $(z_\tau^{(\varepsilon,t,x)})_{0\le\tau\le T}$, it follows that $(D_s z^{(\varepsilon)}(\tau, X_\tau^{(t,x)})))_{s\le\tau\le T}$ is a version of $(D_s z_\tau^{(\varepsilon,t,x)})_{s\le\tau\le T}$, and therefore that

$$
\big(\sigma^{-1}(s, X_s^{(t,x)})\nabla X_s^{(t,x)}\big)^\top D_s z_\tau^{(\varepsilon,t,x)})_{s\le\tau\le T} = ((\nabla X_\tau^{(t,x)})^\top \nabla_x z^{(\varepsilon)}(\tau, X_\tau^{(t,x)}))_{s\le\tau\le T}
$$

is a version of $(\big(\sigma^{-1}(s, X_s^{(t,x)})\nabla X_s^{(t,x)}\big)^\top D_s z_\tau^{(\varepsilon,t,x)})_{s\le\tau\le T}$ for all $s \in [0, T]$.

We now combine the BSDE arguments and the functional arguments from above. Thanks to the intermediate version $(\big(\sigma^{-1}(s, X_s^{(t,x)})\nabla X_s^{(t,x)}\big)^\top D_s z_\tau^{(\varepsilon,t,x)})_{0\le\tau\le T}$, it follows that

$$
((\nabla X_\tau^{(t,x)})^\top \nabla_x z^{(\varepsilon)}(\tau, X_\tau^{(t,x)}))_{0\le\tau\le T}
$$

is a version of $(\nabla z_\tau^{(\varepsilon,t,x)})_{0\le\tau\le T}$.

Step 4. Proving $z^{(\varepsilon)}(t, \cdot)$ is Lipschitz continuous. Fix $s \in [t, T)$. Using the representation (4.4) of $z^{(\varepsilon,t,x)}$, it follows that

$$
\begin{aligned}
\|z_s^{(\varepsilon,t,x_1)} - z_s^{(\varepsilon,t,x_2)}\|_2 \le \|&\mathbb{E}_s[\int_s^T F(r, X_r^{(t,x_1)}, y_r^{(\varepsilon,t,x_1)}, z_r^{(\varepsilon,t,x_1)}) H_r^{(t,x_1,s)} dr] \\
&- \mathbb{E}_s[\int_s^T F(r, X_r^{(t,x_2)}, y_r^{(\varepsilon,t,x_2)}, z_r^{(\varepsilon,t,x_2)}) H_r^{(t,x_1,s)} dr]\|_2 \\
&+ \|\mathbb{E}_s[\int_s^T F(r, X_r^{(t,x_2)}, y_r^{(\varepsilon,t,x_2)}, z_r^{(\varepsilon,t,x_2)})(H_r^{(t,x_1,s)} - H_r^{(t,x_2,s)}) dr]\|_2 \\
=:& \mathcal{A}_1 + \mathcal{A}_2.
\end{aligned}
$$

We start with an estimate for $\mathcal{A}_2$. Using the Cauchy-Schwarz inequality, it follows that

$$\mathcal{A}_2 \leq \int_s^T \|F(r, X_r^{(t,x_2)}, y_r^{(\varepsilon,t,x_2)}, z_r^{(\varepsilon,t,x_2)})\|_4 \|H_r^{(t,x_1,s)} - H_r^{(t,x_2,s)}\|_4 dr \tag{4.8}$$

▶ **Bounding** $\|H_r^{(t,x_1,s)} - H_r^{(t,x_2,s)}\|_4$**.** Using the same techniques as in the proof of Lemma 2.10, one shows that

$$\begin{aligned}\|H_r^{(t,x_1,s)} - H_r^{(t,x_2,s)}\|_4 \leq C_4 &\frac{\|\sigma^{-1}(s, X_s^{(t,x_1)}) - \sigma^{-1}(s, X_s^{(t,x_2)})\|_{\mathcal{S}^8} \mathbb{E}[\sup_{s\leq u\leq T} |D_s X_u^{(t,x_1)}|^8]^{1/8}}{\sqrt{r-s}} \\ &+ C_4 \frac{\|\sigma^{-1}\|_\infty \mathbb{E}[\sup_{s\leq u\leq T} |D_s X_u^{(t,x_1)} - D_s X_u^{(t,x_2)}|^8]^{1/8}}{\sqrt{r-t}}.\end{aligned} \tag{4.9}$$

where C_4 is the constant coming from the BDG inequality. Thanks to [RY99, Theorem IX.2.4], we have that

$$\|X_s^{(t,x_1)} - X_s^{(t,x_2)}\|_{\mathcal{S}^8} \leq C|x_1 - x_2| \tag{4.10}$$

The function $\sigma^{-1}(t, \cdot)$ is Lipschitz continuous uniformly in t with Lipschitz constant as given in Lemma 4.1 for all $s \in [t, T)$. Moreover, Lemma 2.5 gives that

$$\mathbb{E}[\sup_{s\leq u\leq T} |D_s X_u^{(t,x_1)}|^8]^{1/8} \leq C \quad \text{and} \quad \mathbb{E}[\sup_{s\leq u\leq T} |D_s X_u^{(t,x_1)} - D_s X_u^{(t,x_2)}|^8]^{1/8} \leq C|x_1 - x_2|.$$

Combining these estimates, it follows that $\|H_r^{(t,x_1,t)} - H_r^{(t,x_2,t)}\|_4 \leq C|x_1 - x_2|/\sqrt{r-t}$.

▶ **Bounding** $\|F(r, X_r^{(t,x_2)}, y_r^{(\varepsilon,t,x_2)}, z_r^{(\varepsilon,t,x_2)})\|_4$**.** We take advantage of the local Lipschitz continuity and boundedness (1.2) of f, and the uniform bounds on u and its partial derivatives from Lemma 2.8, in order to show that

$$\begin{aligned}|F(r, X_r^{(t,x_2)}, 0, 0)| &\leq |f(r, X_r^{(t,x_2)}, 0, 0)| + L_f \frac{|u(r, X_r^{(t,x_2)})| + \|\sigma\|_\infty |\nabla_x u(r, X_r^{(t,x_2)})|}{(T-r)^{(1-\theta_L)/2}} \\ &\leq \frac{C_f}{(T-r)^{1-\theta_c}} + \frac{CB_r(\Phi)}{(T-r)^{1-\theta_L/2}} \leq C \frac{B_r(\Phi)}{(T-r)^{1-\theta_c \wedge \frac{\theta_L}{2}}}\end{aligned} \tag{4.11}$$

where

$$B_r(\Phi) := \begin{cases} \sqrt{T-t} + C & \text{if } \Phi \text{ is constant,} \\ \mathbb{E}_r[|\Phi(X_T^{(t,x_2)}) - \mathbb{E}_r[\Phi(X_T^{(t,x_2)})]|^2]^{1/2} + C & \text{else.} \end{cases}$$

Without loss of generality, we will consider the setting where Φ is not constant, because, for constant Φ, the arguments will be analogous to the arguments under $(\mathbf{A_{h\Phi}})$ with $\theta_\Phi \equiv 1$. It follows from the triangle inequality, the local Lipschitz continuity (1.2) and the inequality (4.11) that

$$|F(r, X_r^{(t,x_2)}, y_r^{(\varepsilon,t,x_2)}, z_r^{(\varepsilon,t,x_2)})| \leq |F(r, X_r^{(t,x_2)}, 0, 0)| + L_f \frac{|y_r^{(\varepsilon,t,x_2)}| + |z_r^{(\varepsilon,t,x_2)}|}{(T-r)^{(1-\theta_L)/2}}$$

But $y_r^{(\varepsilon,t,x_2)}$ and $z_r^{(\varepsilon,t,x_2)}$ are bounded in $\mathbf{L}_4$: applying Proposition 2.12 with $(Y_1, Z_1) = (0, 0)$ and $(Y_2, Z_2) = (y^{(\varepsilon,t,x_2)}, z^{(\varepsilon,t,x_2)})$, combined with inequality (4.11) and Lemma C.2 to obtain that

$$\begin{aligned}|y_r^{(\varepsilon,t,x_2)}| &\leq C \int_r^{T-\varepsilon} \mathbb{E}_r[|F(u, X_u^{(t,x_2)}, 0, 0)|^2]^{1/2} du \leq CB_r(\Phi)(T-r)^{\theta_c \wedge \theta_L/2}, \\ |z_r^{(\varepsilon,t,x_2)}| &\leq C \int_r^{T-\varepsilon} \mathbb{E}_r[|F(u, X_u^{(t,x_2)}, 0, 0)|^2]^{1/2} (u-r)^{-1/2} du \leq CB_r(\Phi)(T-r)^{\theta_c \wedge \frac{\theta_L}{2} - \frac{1}{2}}\end{aligned} \tag{4.12}$$

for all $r \in [t, T)$. Therefore, $\|F(r, X_r^{(t,x_2)}, y_r^{(\varepsilon,t,x_2)}, z_r^{(\varepsilon,t,x_2)})\|_4$ is bounded above by $C(T-r)^{\theta_c \wedge \frac{\theta_L}{2}-1}$. Now, both $(\mathbf{A_{b\Phi}})$ and $(\mathbf{A_{h\Phi}})$imply that $\|B_r(\Phi)\|_4$ is bounded above uniformly in r by C. Substituting this and the bound on $\|H_r^{(t,x_1,t)} - H_r^{(t,x_2,t)}\|_4$ into (4.8)

$$\mathcal{A}_2 \leq CB_4(\Phi)|x_1 - x_2| \int_s^T \frac{dr}{(T-r)^{1-\theta_c \wedge \frac{\theta_L}{2}} \sqrt{r-s}} \leq \frac{CB_4(\Phi)|x_1 - x_2|}{(T-s)^{\frac{1}{2}-\theta_c \wedge \frac{\theta_L}{2}}}$$

Now, we estimate $\mathcal{A}_1$. Using Corollary 2.11 (with $H_r^{(t,x_1,s)}$ in the place of H_r^s), it follows that

$$\mathcal{A}_1 \leq \| \int_s^{T-\varepsilon} (\mathbb{E}_s[|F(r, X_r^{(t,x_1)}, y_r^{(\varepsilon,t,x_1)}, z_r^{(\varepsilon,t,x_1)}) - F(r, X_r^{(t,x_2)}, y_r^{(\varepsilon,t,x_2)}, z_r^{(\varepsilon,t,x_2)})|^2])^{1/2} (\mathbb{E}_s[|H_r^{(t,x_1,s)}|^2])^{1/2} dr\|_2$$

Analogously to Lemma 2.10, $(\mathbb{E}_s[|H_r^{(t,x_1,s)}|^2])^{1/2} \leq C_M (r-s)^{-1/2}$, therefore

$$\begin{aligned}
\mathcal{A}_1 &\leq C\| \int_s^{T-\varepsilon} \frac{(\mathbb{E}_s[|F(r, X_r^{(t,x_1)}, y_r^{(\varepsilon,t,x_1)}, z_r^{(\varepsilon,t,x_1)}) - F(r, X_r^{(t,x_2)}, y_r^{(\varepsilon,t,x_2)}, z_r^{(\varepsilon,t,x_2)})|^2])^{1/2}}{\sqrt{r-s}} dr\|_2 \\
&\leq C \int_s^{T-\varepsilon} \frac{\|(\mathbb{E}_s[|F(r, X_r^{(t,x_1)}, y_r^{(\varepsilon,t,x_1)}, z_r^{(\varepsilon,t,x_1)}) - F(r, X_r^{(t,x_2)}, y_r^{(\varepsilon,t,x_2)}, z_r^{(\varepsilon,t,x_2)})|^2])^{1/2}\|_2}{\sqrt{r-s}} dr \\
&= C \int_s^{T-\varepsilon} \frac{\|F(r, X_r^{(t,x_1)}, y_r^{(\varepsilon,t,x_1)}, z_r^{(\varepsilon,t,x_1)}) - F(r, X_r^{(t,x_2)}, y_r^{(\varepsilon,t,x_2)}, z_r^{(\varepsilon,t,x_2)})\|_2}{\sqrt{r-s}} dr
\end{aligned}$$

where we have used Minkowski's inequality to take the norm $\|\cdot\|_2$ into the Lebesgue integral. By applying the Lipschitz continuity of $f^{(\varepsilon)}(r, \cdot)$, $\mathcal{A}_1$ is bounded by

$$\begin{aligned}
&C \int_s^{T-\varepsilon} \frac{\|X_r^{(t,x_1)} - X_r^{(t,x_2)}\|_2}{(T-r)^{1-\theta_X/2}\sqrt{r-t}} dr \\
&\quad + C \int_s^{T-\varepsilon} \frac{\|\sigma\|_\infty \|u(r, X_r^{(t,x_1)}) - u(r, X_r^{(t,x_2)})\|_2 + \|\nabla_x \sigma\|_\infty \|\nabla_x u(r, X_r^{(t,x_1)}) - \nabla_x u(r, X_r^{(t,x_2)})\|_2}{(T-r)^{(1-\theta_L)/2}\sqrt{r-t}} dr \\
&\quad + C \int_s^{T-\varepsilon} \frac{\|y_r^{(\varepsilon,t,x_1)} - y_r^{(\varepsilon,t,x_2)}\|_2 + \|z_r^{(\varepsilon,t,x_1)} - z_r^{(\varepsilon,t,x_2)}\|_2}{(T-r)^{(1-\theta_L)/2}\sqrt{r-t}} dr
\end{aligned}$$

Using the differentiability of $u(s, \cdot)$, it follows that

$$\begin{aligned}
&\|u(r, X_r^{(t,x_1)}) - u(r, X_r^{(t,x_2)})\|_2 \leq \|\mathcal{R}(u, r, X_r^{(t,x_1)}, X_r^{(t,x_2)})\|_2, \\
&\|\nabla_x u(r, X_r^{(t,x_1)}) - \nabla_x u(r, X_r^{(t,x_2)})\|_2 \leq \|\mathcal{R}(\nabla_x u, r, X_s^{(t,x_1)}, X_s^{(t,x_2)})\|_2
\end{aligned}$$

for all $r \in [t, T)$, where, for a differentiable function g, $\mathcal{R}(g, r, x, x')$ is the remainder from the first order Taylor expansion of $g(r, x) - g(r, x')$: in the case of g taking values in $\mathbb{R}$, this is equal to

$$\mathcal{R}(g, r, x, x') = \{\int_0^1 \nabla_x g(\delta x + (1-\delta)x') d\delta\}(x - x'); \tag{4.13}$$

in the multidimensional case, the expansion (4.13) is defined component-wise. Denote by $\mathcal{R}(r)$ the sum of the normed residuals $\|\mathcal{R}(u, r, X_r^{(t,x_1)}, X_r^{(t,x_2)})\|_2 + \|\mathcal{R}(\nabla_x u, r, X_s^{(t,x_1)}, X_s^{(t,x_2)})\|_2$. Therefore, using the notation $\Theta_r := \|y_r^{(\varepsilon,t,x_1)} - y_r^{(\varepsilon,t,x_2)}\|_2 + \|z_r^{(\varepsilon,t,x_1)} - z_r^{(\varepsilon,t,x_2)}\|_2$, the final bound on $\mathcal{A}_1$ is

$$\mathcal{A}_1 \leq C \int_s^{T-\varepsilon} \frac{\|X_r^{(t,x_1)} - X_r^{(t,x_2)}\|_2 dr}{(T-r)^{1-\theta_X/2}\sqrt{r-t}} + C \int_s^{T-\varepsilon} \frac{\mathcal{R}(r) dr}{(T-r)^{(1-\theta_L)/2}\sqrt{r-t}} + C \int_s^{T-\varepsilon} \frac{\Theta_r dr}{(T-r)^{(1-\theta_L)/2}\sqrt{r-t}}.$$

It follows from Lemma 2.8 that

$$\mathcal{R}(r) \leq \begin{cases} C\|\Phi\|_\infty \|X_r^{(t,x_1)} - X_r^{(t,x_2)}\|_2 (T-r)^{-1} \leq C\|\Phi\|_\infty |x_1 - x_2| (T-r)^{-1} & \text{under } (\mathbf{A_{b\Phi}}), \\ \\ CK_\Phi \|X_r^{(t,x_1)} - X_r^{(t,x_2)}\|_2 (T-r)^{\frac{\theta_\Phi}{2}-1} \leq CK_\Phi |x_1 - x_2| (T-r)^{\frac{\theta_\Phi}{2}-1} & \text{under } (\mathbf{A_{h\Phi}}), \end{cases}$$

The bound $\|X_r^{(t,x_1)} - X_r^{(t,x_2)}\|_2 \leq C|x_1 - x_2|$ is obtained from [RY99, Theorem IX.2.4], which also implies that

$$\int_s^{T-\varepsilon} \frac{\|X_r^{(t,x_1)} - X_r^{(t,x_2)}\|_2}{(T-r)^{1-\theta_X/2}\sqrt{r-t}} dr \leq \int_s^{T-\varepsilon} \frac{C|x_1 - x_2|}{(T-r)^{1-\theta_X/2}\sqrt{r-t}} dr.$$

It is clear from the bounds above that the integral in $\mathcal{R}(r)$ in the bound of $\mathcal{A}_1$ dominates the upper bound on $\mathcal{A}_2$, and also the integral

$$\int_s^{T-\varepsilon} \frac{\|X_r^{(t,x_1)} - X_r^{(t,x_2)}\|_2}{(T-r)^{1-\theta_X/2}\sqrt{r-t}} dr.$$

Therefore,

$$\|z_s^{(\varepsilon,t,x_1)} - z_s^{(\varepsilon,t,x_2)}\|_2 \leq C \int_s^{T-\varepsilon} \frac{\mathcal{R}(r)dr}{(T-r)^{(1-\theta_L)/2}\sqrt{r-t}} + C\int_s^{T-\varepsilon} \frac{\Theta_r}{(T-r)^{(1-\theta_L)/2}\sqrt{r-t}} dr.$$

Since $y_s^{(\varepsilon,t,x)} = \mathbb{E}_s[\int_s^T F(r, X_r^{(t,x)}, y_r^{(\varepsilon,t,x)}, z_r^{(\varepsilon,t,x)})dr]$ similar estimates yield

$$\Theta_s \leq C\int_s^{T-\varepsilon} \frac{\mathcal{R}(r)dr}{(T-r)^{(1-\theta_L)/2}\sqrt{r-s}} + C\int_s^{T-\varepsilon} \frac{\Theta_r}{(T-r)^{(1-\theta_L)/2}\sqrt{r-s}} dr$$

for all $s \in [t, T-\varepsilon)$. Let $(\mathbf{A_{b\Phi}})$ be in force. Applying Lemma C.3 with

$$w_r := C\|\Phi\|_\infty |x_1 - x_2| \int_r^{T-\varepsilon} (T-u)^{(\theta_L-3)/2}(u-r)^{-1/2} du$$

and $u_r := \Theta_r$, it follows that

$$\begin{aligned}
\Theta_s &\leq C\|\Phi\|_\infty |x_1 - x_2| \int_s^{T-\varepsilon} \frac{dr}{(T-r)^{(3-\theta_L)/2}\sqrt{r-s}} \\
&\quad + C\|\Phi\|_\infty |x_1 - x_2| \int_s^{T-\varepsilon} \frac{\int_s^u (u-r)^{\theta_L/2-1}(r-s)^{-1/2}dr}{(T-u)^{(3-\theta_L)/2}} du + C\int_s^{T-\varepsilon} \frac{\Theta_r}{(T-r)^{(1-\theta_L)/2}} dr \\
&\leq C\|\Phi\|_\infty |x_1 - x_2| \int_s^{T-\varepsilon} \frac{dr}{(T-r)^{(3-\theta_L)/2}\sqrt{r-s}} + C\int_s^{T-\varepsilon} \frac{\Theta_r}{(T-r)^{(1-\theta_L)/2}} dr,
\end{aligned}$$

where we have used Lemma C.2 to bound the integral $\int_s^u (u-r)^{\theta_L/2-1}(r-s)^{-1/2}dr$. Then, applying Lemma C.4 to bound the integral $\int_s^{T-\varepsilon} \Theta_r (T-r)^{(\theta_L-1)/2} dr$, final bound on $\|z_s^{(\varepsilon,t,x_1)} - z_s^{(\varepsilon,t,x_2)}\|_2$ for all $t \in [0,T)$, $(x_1, x_2) \in (\mathbb{R}^d)^2$, and $s \in [t, T)$ is

$$\begin{aligned}
\|z_s^{(\varepsilon,t,x_1)} - z_s^{(\varepsilon,t,x_2)}\|_2 &\leq C\|\Phi\|_\infty |x_1 - x_2| \int_s^{T-\varepsilon} \frac{dr}{(T-r)^{(3-\theta_L)/2}\sqrt{r-s}} \\
&\quad + C\|\Phi\|_\infty |x_1 - x_2| \int_s^{T-\varepsilon} \frac{\int_r^{T-\varepsilon} (T-u)^{(\theta_L-3)/2}(u-r)^{-1/2} du}{(T-r)^{(1-\theta_L)/2}} dr
\end{aligned}$$

and application of Lemma C.2 yields the final upper bound $C\phi(t, \varepsilon, \theta_L)|x_1 - x_2|$. Therefore, setting $t = s$ yields that the function $z^{(\varepsilon)}(t, \cdot)$ is Lipschitz continuous with Lipschitz constant $C\phi(t, \varepsilon, \theta_L)$, as required. The proof under $(\mathbf{A_{h\Phi}})$ is analogous. □

We now come to the regularity result advertised at the beginning of this section; this result is not used in the remainder of this paper, but may hold some interest for other works.

Corollary 4.3. *Let* $(\mathbf{A_{h\Phi}})$ *and* $(\mathbf{A_{\partial f}})$ *be in force, and let* $\theta_L+\theta_\Phi \geq 1$*. Then there exists a function* $z : [0,T)\times\mathbb{R}^d \to (\mathbb{R}^q)^\top$ *such that* $\nabla_x u(s,X_s)\sigma(s,X_s)+z(s,X_s)$ *is a version of* Z_s*. Moreover, recalling the function* $\phi(t,\varepsilon,\theta_L,\theta_\Phi)$ *from (4.2) for all* $t\in[0,T)$*,* $x\mapsto z(t,x)$ *is Lipschitz continuous with Lipschitz constant equal to*

$$\lim_{\varepsilon\to 0}\phi(t,\varepsilon,\theta_L,\theta_\Phi)\leq \frac{C}{(T-t)^{1-(\theta_L+\theta_\Phi)/2}}$$

for some finite constant C *depending only on* K_Φ*,* L_f*, the bounds on* b *and* σ *and their partial derivatives,* $\bar\beta$*,* C_M*,* θ_L*,* θ_c*,* C_f*, and* T*.*

Proof. Let $(Y^{(t,x)},Z^{(t,x)})$ be the solution of

$$Y_s^{(t,x)}=\Phi(X_T^{(t,x)})+\int_s^T f(\tau,X_\tau^{(t,x)},Y_\tau^{(t,x)},Z_\tau^{(t,x)})d\tau-\int_s^T Z_\tau^{(t,x)}dW_\tau,$$

and set

$$z(t,x):=\mathbb{E}_t[\int_s^T f(\tau,X_\tau^{(t,x)},Y_\tau^{(t,x)},Z_\tau^{(t,x)})H_\tau^{(t,x,s)}d\tau]$$

for

$$H_r^{(t,x,s)}:=\frac{\mathbf{1}_{(t,T]}(s)}{r-s}(\int_s^r \sigma^{-1}(r,X_r^{(t,x)})D_sX_r^{(t,x)})^\top dW_r)^\top$$

Recall the function $z^{(\varepsilon)}:[0,T)\times\mathbb{R}^d\to(\mathbb{R}^q)^\top$ from (4.4). One shows $z(t,x)=\lim_{\varepsilon\to0}z^{(\varepsilon)}(t,x)$ by mimicking the proof of Theorem 2.16. Since Z is the limit of $Z^{(\varepsilon)}$ as $\varepsilon\to0$ in $\mathcal{H}^2$, and $z^{(\varepsilon)}(s,X_s)$ is a version of $Z_s^{(\varepsilon)}-\nabla_x u(s,X_s)\sigma(s,X_s)$, it follows that $z(s,X_s)$ is a version of $Z_s-\nabla_x u(s,X_s)\sigma(s,X_s)$, as required. Finally, to prove the Lipschitz continuity of $z(t,\cdot)$, we observe that, for $\theta_L+\theta_\Phi\geq1$,

$$\lim_{\varepsilon\to0}\phi(t,\varepsilon,\theta_L,\theta_\Phi)=K_\Phi\int_t^T\frac{dr}{(T-r)^{(3-\theta_L-\theta_\Phi)/2}\sqrt{r-t}}\leq\frac{C}{(T-t)^{1-(\theta_L+\theta_\Phi)/2}}$$

thanks to Lemma C.2, and proceed as in **Step 4** of the proof of Proposition 4.2 (with $z(t,\cdot)$ in the place of $z^{(\varepsilon)}(t,\cdot)$); the upper bound on the limit $\lim_{\varepsilon\to0}\phi(t,\varepsilon,\theta_L,\theta_\Phi)$ comes from Lemma C.2. □

In order to make use of Proposition 4.2, it is is necessary to approximate Z by an intermediate process Z_M which satisfies the hypotheses of Proposition 4.2.

Lemma 4.4. *Assume that* $(\mathbf{A_{exp\Phi}})$ *is in force. Recall the BSDE* (Y_M,Z_M) *defined in Corollary 2.15. Take the version of* Z_M *given by Theorem 2.16. For* $M=(3\ln(N))^{1/4}$ *and* $R(M)$ *equal to* $3L_fe^{M^2/2}$*, there is a finite constant* C *depending only on* L_f*,* C_M*,* θ_L*,* C_ξ *and* T*, but not on* N*, such that for all* $N\geq1$

$$\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\|Z_s-\tilde Z_{t_i}\|_2^2ds\leq C\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\|Z_{M,s}-\tilde Z_{M,t_i}\|_2^2ds+CN^{-1}$$

where $\tilde Z_{t_i}:=\frac{1}{\Delta_i}\mathbb{E}_i\big[\int_{t_i}^{t_{i+1}}Z_tdt\big]$ *and* $\tilde Z_{M,t_i}:=\frac{1}{\Delta_i}\mathbb{E}_i\big[\int_{t_i}^{t_{i+1}}Z_{M,t}dt\big]$*.*

Proof. In what follows, C may change from line to line. Using Cauchy's inequality and the orthogonality of the projections,

$$\frac12\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\|Z_s-\tilde Z_{t_i}\|_2^2ds\leq\int_0^T\|Z_s-Z_{M,s}\|_2^2ds+\sum_{i=0}^{N-1}\|\tilde Z_{t_i}-\tilde Z_{M,t_i}\|_2^2\Delta_i+\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\|Z_{M,s}-\tilde Z_{M,t_i}\|_2^2ds.$$

From Jensen's inequality, it follows that $\sum_{i=0}^{N-1}\|\tilde{Z}_{t_i}-\tilde{Z}_{M,t_i}\|_2^2\Delta_i \le \int_0^T \|Z_s-\bar{Z}_{M,s}\|_2^2 ds$. Proposition 2.12 with $(Y_1, Z_1) := (Y_M, Z_M)$ and $(Y_2, Z_2) := (Y, Z)$ yields that, for any $s \in [0, T)$,

$$\int_0^T \|Z_t - Z_{M,t}\|_2^2 dt \le C\|\Phi(X_T) - \Phi_M(X_T)\|_2^2 + C\int_0^T \|f(t, X_t, Y_{M,t}, Z_{M,t}) - f_M(t, X_t, Y_{M,t}, Z_{M,t})\|_2^2 dt, \tag{4.14}$$

It follows from Markov's exponential inequality and $(\mathbf{A_{exp\Phi}})$ that

$$\|\Phi(X_T) - \Phi_M(X_T)\|_2^2 = \int_{M^2}^\infty \mathbb{P}(|\Phi(X_T)|^2 \ge x)dx \le C_\xi \int_{M^2}^\infty e^{-\sqrt{x}}dx = 2C_\xi(1+M^4)e^{-M^4} \le CN^{-2}. \tag{4.15}$$

The last inequality is obtained by substituting the value of M. On the other hand, the basic properties of the mollifier in Definition 1.1 yields

$$\begin{aligned}
&|f(t, X_t, Y_{M,t}, Z_{M,t}) - f_M(t, X_t, Y_{M,t}, Z_{M,t})| \\
&\le \int_{\mathbb{R}^d\times\mathbb{R}\times(\mathbb{R}^q)^\top} |f(t, X_t, Y_{M,t}, Z_{M,t}) - f(t, X_t - x, Y_{M,t} - y, Z_{M,t} - z)|\phi_{R(M)}(x, y, z)d(x, y, z) \\
&\le \frac{L_f}{(T-t)^{(1-\theta_L)/2}} \int_{\{|x|^2+|y|^2+|z|^2\le R(M)^{-2}\}} (|x| + |y| + |z|)\phi_{R(M)}(x, y, z)d(x, y, z) \le \frac{3L_f}{(T-t)^{(1-\theta_L)/2}R(M)}
\end{aligned} \tag{4.16}$$

Substituting the value of $R(M)$ then gives $\|f(t, X_t, \bar{Y}_{M,t}, \bar{Z}_{M,t}) - f_M(t, X_t, \bar{Y}_{M,t}, \bar{Z}_{M,t}\|_2 \le (T - t)^{(\theta_L-1)/2}N^{-1/2}$ for all $t \in [0, T)$. Substituting (4.15) and (4.16) into (4.14) Lemma C.2 then yields

$$\int_0^T \|Z_s - Z_{M,s}\|_2^2 ds \le CN^{-1} + CN^{-2}\sum_{i=0}^{N-1}\frac{\Delta_i}{T-t_i},$$

The sum on the right hand side above is bounded by $1 + \int_0^{t_{N-1}}(T-t)^{-1}dt = 1 + C\ln(N)$, whence the proof is complete. $\square$

We now provide an extension to Theorem 3.3 under $(\mathbf{A_{h\Phi}})$ with the aid of Proposition 4.2.

Theorem 4.5. *Let $(\mathbf{A_{h\Phi}})$ be in force and $0 < \beta < (2\gamma)\wedge(\alpha\wedge\theta_L)$. There is a constant C depending only on L_f, C_M, θ_L, θ_c, β, C_f, K_Φ and T, but not on N, such that for all $N \ge 1$,*

$$\mathcal{E}(N) \le CN^{-1}\mathbf{1}_{[1,3]}(\theta_\Phi + \beta + 2\gamma) + CN^{-2\gamma}\mathbf{1}_{(0,1)}(\theta_\Phi + \beta + 2\gamma) \tag{4.17}$$

for $\gamma := (\theta_c \wedge \frac{\alpha}{2} + \frac{\theta_L}{2}) \wedge \theta_c$.

Proof. In what follows, C may change from line to line. To start with, we assume that $(\mathbf{A_{\partial f}})$ and $(\mathbf{A_{b\Phi}})$ are in force. Recall (3.2). From the bounds $\|a_r^{(\varepsilon)}\|_2 \le C(T-r)^{(\alpha+\theta_L-3)/2}$ in the proof on Lemma 2.9 the first sum $\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}(\int_{t_i}^t \|a_r^{(\varepsilon)}\|_2 dr)^2 dt$ is bounded above by

$$\begin{aligned}
&C\sum_{i=0}^{N-2}\int_{t_i}^{t_{i+1}}\frac{\left(\int_{t_i}^t \frac{dr}{(t-r)^{(1-\theta_L)/2}}\right)^2}{(T-t)^{2-\alpha}}dt + C\int_{t_{N-1}}^T \frac{\left(\int_{t_{N-1}}^t \frac{dr}{(t-r)^{1-\theta_L/2}}\right)^2}{(T-t)^{1-\alpha}}dt \\
&\le C\sum_{i=0}^{N-2}\int_{t_i}^{t_{i+1}}\frac{(T-t)^{\beta-1}(t-t_i)^{1+\theta_L}}{(T-t)^{1+\beta-\alpha}}dt + C\int_{t_{N-1}}^T \frac{(t-t_{N-1})^{\theta_L}}{(T-t)^{1-\alpha}}dt \\
&\le C\sum_{i=0}^{N-2}\frac{\Delta_i^{1+\theta_L}}{(T-t_{i+1})^{1-\beta}}\int_{t_i}^{t_{i+1}}\frac{dt}{(T-t)^{1+\beta-\alpha}} + C\Delta_{N-1}^{\theta_L+\alpha}.
\end{aligned} \tag{4.18}$$

Using (B.2) from Lemma B.1, $\Delta_i \le C\Delta_{i+1}$ for $i < N-1$, which, combined with (B.1), yields

$$\max_{0\le i\le N-2} \frac{\Delta_i^{1+\theta_L}}{(T-t_{i+1})^{1-\beta}} \le C \max_{0\le i\le N-1} \frac{\Delta_i^{1+\theta_L}}{(T-t_i)^{1-\beta}} < CN^{-1-\theta_L}.$$

Additionally, $\beta < \alpha$ implies that $\Delta_{N-1}^{\alpha+\theta_L} = CN^{-(\alpha+\theta_L)/\beta} \le CN^{-1}$. Substituting these results into (4.18) gives

$$\sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} \Big(\int_{t_i}^{t} \|a_r^{(\varepsilon)}\|_2 dr\Big)^2 dt \le CN^{-1}. \tag{4.19}$$

The refined estimates – $\|V_r^{(\varepsilon)}\|_2 \le C\phi(r,\varepsilon,\theta_L,\theta_\Phi)$ for all $r\in[0,T)$ – from Proposition 4.2 are used to bound $\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}(\int_{t_i}^{t}\|V_r^{(\varepsilon)}\|_2 dr)^2 dt$ from above by $\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\Big(\int_{t_i}^{t}\phi(r,\varepsilon,\theta_L,\theta_\Phi)dr\Big)^2 dt$, which itself is bounded above by

$$\begin{aligned}
\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\Big(\int_{t_i}^{t}\phi(r,\varepsilon,\theta_L,\theta_\Phi)dr\Big)^2 dt &= \sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\Big(\int_{t_i}^{t}\Big\{K_\Phi\int_r^{T-\varepsilon}\frac{du}{(T-u)^{\frac{3-\theta_\Phi-\theta_L}{2}}\sqrt{u-r}}\Big\}dr\Big)^2 dt\\
&= K_\Phi^2\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\Big(\int_{t_i}^{t}\Big\{\int_r^{T-\varepsilon}\frac{(T-u)^{(\beta-1)/2}du}{(T-u)^{\frac{2+\beta-\theta_\Phi-\theta_L}{2}}\sqrt{u-r}}\Big\}dr\Big)^2 dt\\
&\le \frac{K_\Phi^2}{\varepsilon^{1-\theta_\Phi-\beta}}\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\Big(\int_{t_i}^{t}\Big\{\int_r^{T-\varepsilon}\frac{du}{(T-u)^{1-(\theta_L-\beta)/2}\sqrt{u-r}}\Big\}dr\Big)^2 dt.
\end{aligned}$$

Now, using Lemma C.2 to obtain an upper bound $(T-r)^{-\big(1-(\theta_L-\beta)\big)/2}$ on the inner integral $\int_r^{T-\varepsilon}(T-u)^{-\big(1-(\theta_L-\beta)/2\big)}(u-r)^{-1/2}du$,

$$\begin{aligned}
\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\Big(\int_{t_i}^{t}\phi(r,\varepsilon,\theta_L,\theta_\Phi)dr\Big)^2 dt &\le \frac{C}{\varepsilon^{1-\theta_\Phi-\beta}}\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\Big(\int_{t_i}^{t}\frac{dr}{(T-r)^{\big(1-(\theta_L-\beta)\big)/2}}\Big)^2 dt\\
&\le \frac{C(\max_{0\le i\le N-1}\Delta_i)^2}{\varepsilon^{1-\theta_\Phi-\beta}}\int_0^T\frac{dr}{(T-t)^{1-(\theta_L-\beta)}}dt \le \frac{CN^{-2}}{\varepsilon^{1-\theta_\Phi-\beta}}
\end{aligned} \tag{4.20}$$

where we have used Jensen's inequality to get

$$\Big(\int_{t_i}^{t}\frac{dr}{(T-r)^{\big(1-(\theta_L-\beta)\big)/2}}\Big)^2 \le C\int_0^T\frac{dr}{(T-t)^{1-(\theta_L-\beta)}}dt,$$

and then (B.1) in Lemma B.1 for the bound $\max_{0\le i\le N-1}\Delta_i \le CN^{-1}$. Substituting (4.19) and (4.20) into (3.2) finally yields

$$\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\|z_s^{(\varepsilon)}-z_{t_i}^{(\varepsilon)}\|_2^2 ds \le CN^{-1} + \frac{CN^{-2}}{\varepsilon^{1-\theta_\Phi-\beta}}. \tag{4.21}$$

Then, using $Z^{(\varepsilon)} = z^{(\varepsilon)} + z$, Lemma 3.2, and $\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\|z_s - z_{t_i}\|_2^2 ds \le CN^{-1}$ as shown in [GM10, Theorem 1.3], it follows that

$$\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\|Z_s - Z_{t_i}\|_2^2 ds \le CN^{-1} + CN^{-2}\varepsilon^{\theta_\Phi+\beta-1} + CN^{-2\gamma/\beta} + C\varepsilon^{2\gamma}. \tag{4.22}$$

For $N\in\{1,2\}$, let $\delta := 0$, and for $N>2$, $\delta := \ln\ln(N)/\ln(N)$. Set $\varepsilon = N^{-(1+\delta)/(2\gamma)}$. Recalling further that $2\gamma<\beta$, this implies that, under $(\mathbf{A_{b\Phi}})$ and $(\mathbf{A_{\partial f}})$,

$$\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\|Z_s - Z_{t_i}\|_2^2 ds \le CN^{-1} + CN^{-2-(\theta_\Phi+\beta-1)(1+\delta_N)/(2\gamma)}.$$

To obtain the general result, recall the BSDE (Y_M, Z_M) from Corollary 2.15. The driver of (Y_M, Z_M) satisfies assumptions $(\mathbf{A}_{\partial \mathbf{f}})$. The proof is complete by taking M equal to $(3\ln(N))^{1/4}$, $R(M)$ equal to $3L_f e^{M^4/2}$, and applying Lemma 4.4. $\square$

5 Convergence rate of the Malliavin weights scheme

In this section, we treat the Malliavin weights scheme

$$\bar{Y}_N^{(N)} := \Phi(X_T), \quad \bar{Y}_i^{(N)} := \mathbb{E}_i[\Phi(X_T) + \sum_{j=i}^{N-1} f(t_j, X_{t_j}, \bar{Y}_{j+1}^{(N)}, \bar{Z}_j^{(N)})(t_{j+1} - t_j)],$$

$$\bar{Z}_i^{(N)} := \mathbb{E}_i[\Phi(X_T)H_N^i + \sum_{j=i+1}^{N-1} f(t_j, X_{t_j}, \bar{Y}_{j+1}^{(N)}, \bar{Z}_j^{(N)})H_j^i(t_{j+1} - t_j)]$$

Recall the Malliavin derivative of the the marginals of the process X in Section 2.2. In the definition of the Malliavin weights scheme (1.5), we use the following discrete-time approximation of the Malliavin weights (2.14):

$$H_j^i := \frac{1}{t_j - t_i} \Big(\sum_{k=i}^{j-1} D_{t_i} X_{t_k} \sigma(t_i, X_{t_i}))^\top \Delta W_k \Big)^\top \tag{5.1}$$

Notice that H_j^i satisfies $\mathbb{E}_i[H_j^i] = 0$ and $\mathbb{E}_i[|H_j^i|^2] \le C_M(t_j - t_i)^{-1}$; the latter property is proved exactly like Lemma 2.10. If the marginals of X and $D_{t_i}X$ are not known explicitly, one can use an SDE scheme to provide approximations, but this is beyond the scope of this work; some work has been done on this in the zero driver case ($f \equiv 0$), in particular we refer the reader to Section 3 (and the sequel) of [GM+05]. In what follows, we use the version of Z given by Theorem 2.16, in other words

$$Z_t = \mathbb{E}_t[\Phi(X_T)H_T^t + \int_t^T f(s, X_s, Y_s, Z_s)H_s^t ds] \quad \text{for all } t \in [0, T) \quad \mathbb{P} - a.s.$$

We start with some preliminary results.

Lemma 5.1. *There is a constant C depending only on the bound on b and it's derivatives, the bound on σ and it's derivatives, $\bar{\beta}$, L_f, θ_L, C_f, θ_c, $K^\alpha(\Phi)$ and T such that, for any $0 \le i < j \le N$,*

$$\|\mathbb{E}_i[\Phi(X_T)(H_{t_j}^{t_i} - H_j^i)]\|_2 \le \frac{CN^{-1/2}}{(T - t_i)^{(1-\alpha)/2}},$$

$$\|\mathbb{E}_i[f_j(X_{t_j}, Y_{t_{j+1}}, Z_{t_j})(H_{t_j}^{t_i} - H_j^i)]\|_2 \le \frac{CN^{-1/2}}{(T - t_j)^{1-\gamma}\sqrt{t_j - t_i}}$$

where $\gamma := (\theta_c \wedge \frac{\alpha}{2} + \frac{\theta_L}{2}) \wedge \theta_c$.

Proof. For any $j > i$ and $t \ge t_j$, define $N_t^{t_i} := \sigma^{-1}(t, X_t)D_{t_i}X_t\sigma(t_i, X_{t_i})$. Using the decomposition

$$\begin{aligned} N_t^{t_i} - N_{t_j}^{t_i} &= \sigma^{-1}(t, X_t)(D_{t_i}X_t - D_{t_i}X_{t_j})\sigma(t_i, X_{t_i}) \\ &\quad + (\sigma^{-1}(t, X_t) - \sigma^{-1}(t_j, X_{t_j}))D_{t_i}X_{t_j}\sigma(t_i, X_{t_i}), \end{aligned}$$

it follows from the boundedness and Lipschitz continuity of σ and σ^{-1} (Lemma 4.1) that for any $j > i$ and $t \in [t_j, t_{j+1}]$,

$$\mathbb{E}_i[|H_{t_j}^{t_i} - H_j^i|^2] = \frac{\sum_{k=i}^{j-1} \int_{t_k}^{t_{k+1}} \mathbb{E}_i[|N_t^{t_i} - N_{t_k}^{t_i}|^2]dt}{(t_j - t_i)^2} \le \frac{C \sum_{k=i}^{j-1} \int_{t_k}^{t_{k+1}} \mathbb{E}_i[|D_{t_i}X_t - D_{t_i}X_{t_k}|^2 + |X_t - X_{t_k}|^2]dt}{(t_j - t_i)^2}.$$

It now follows from Lemma 2.5 the usual bound $\mathbb{E}_i[|X_t - X_{t_j}|^2] \le C(t-t_j)$ and Lemma B.1 that

$$\mathbb{E}_i[|H^{t_i}_{t_j} - H^i_j|^2] \le C \max_k \Delta_k (t_j - t_i)^{-1} \le CN^{-1}(t_j - t_i)^{-1}. \tag{5.2}$$

Since $\mathbb{E}_i[H^{t_i}_{t_j} - H^i_j] = 0$, it follows that

$$\|\mathbb{E}_i[\Phi(X_T)(H^{t_i}_{t_j} - H^i_j)]\|_2 = \|\mathbb{E}_i[\{\Phi(X_T) - \mathbb{E}_i[\Phi(X_T)]\}(H^{t_i}_{t_j} - H^i_j)]\|_2.$$

The upper bound

$$\|\mathbb{E}_i[\{\Phi(X_T) - \mathbb{E}_i[\Phi(X_T)]\}(H^{t_i}_{t_j} - H^i_j)]\|_2 \le \|(\mathbb{E}_i[|\{\Phi(X_T) - \mathbb{E}_i[\Phi(X_T)]|^2])^{1/2}(\mathbb{E}_i[|H^{t_i}_{t_j} - H^i_j)|^2])^{1/2}\|_2$$

follows from the conditional Cauchy-Schwarz inequality (Corollary 2.11). Therefore, (5.2) and $\|\Phi(X_T) - \mathbb{E}_i[\Phi(X_T)]\|_2 \le K^\alpha(\Phi)(T-t)^{\alpha/2}$ (from $(\mathbf{A_\Phi})$) together imply that

$$\|\mathbb{E}_i[\Phi(X_T)(H^{t_i}_{t_j} - H^i_j)]\|_2 \le \frac{CN^{-1/2}}{(T-t_i)^{(1-\alpha)/2}}$$

as required. The upper bound on $\|\mathbb{E}_i[f_j(X_{t_j}, Y_{t_{j+1}}, Z_{t_j})(H^{t_i}_{t_j} - H^i_j)]\|_2$ follows from the Cauchy-Schwarz inequality (Corollary 2.11), i.e.

$$\|\mathbb{E}_i[f_j(X_{t_j}, Y_{t_{j+1}}, Z_{t_j})(H^{t_i}_{t_j} - H^i_j)]\|_2 \le \|(\mathbb{E}_i[|f_j(X_{t_j}, Y_{t_{j+1}}, Z_{t_j})|^2])^{1/2}(\mathbb{E}_i[|H^{t_i}_{t_j} - H^i_j)|^2])^{1/2}\|_2;$$

from here, one applies the estimate (5.2) and the fact that, similarly to (2.22), $\|f_j(X_{t_j}, Y_{t_{j+1}}, Z_{t_j})\|_2 \le C(T-t_j)^{\gamma-1}$. ☐

Lemma 5.2. *For all $t_i, t_j \in \pi$ such that $t_i \le t_j$ and $r \in [t_j, T]$,*

$$\mathbb{E}_i[f(r, X_r, Y_r, Z_r)H^{t_i}_r] = \mathbb{E}_i[f(r, X_r, Y_r, Z_r)H^{t_i}_{t_j}]. \tag{5.3}$$

Moreover,

$$\begin{aligned}\mathbb{E}_i[\int_{t_i}^T f(r, X_r, Y_r, Z_r)H^{t_i}_r dr] &= \mathbb{E}_i[\sum_{j=i+1}^{N-1} f_j(X_{t_j}, Y_{t_{j+1}}, Z_{t_j})H^{t_i}_{t_j}\Delta_j] + \mathbb{E}_i[\int_{t_i}^{t_{i+1}} f(r, X_r, Y_r, Z_r)H^{t_i}_r dr] \\ &+ \mathbb{E}_i[\sum_{j=i+1}^{N-1}\int_{t_j}^{t_{j+1}} (f(r, X_r, Y_r, Z_r) - f_j(X_{t_j}, Y_{t_{j+1}}, Z_{t_j}))H^{t_i}_{t_j} dr]. \end{aligned} \tag{5.4}$$

Proof. First let $(\mathbf{A_{\partial f}})$ be in force and recall, as argued in the proof of Theorem 2.16, that the BSDE solved by $(y^{(\varepsilon)}, z^{(\varepsilon)})$ in Definition 2.7 satisfies the conditions of [MZ02, Theorem 4.2]. A key element of the proof of that Theorem is to show that, for almost all $v \in [0, r)$,

$$\begin{aligned} D_v f^{(\varepsilon)}(\Theta_r)\sigma^{-1}(v, X_v)\nabla X_v = \nabla_x f^{(\varepsilon)}(\Theta_r)\nabla X_r + \nabla_y f^{(\varepsilon)}(\Theta_r)(u(r, X_r)\nabla X_r + \nabla y^{(\varepsilon)}_r) \\ + \nabla_z f^{(\varepsilon)}(\Theta_r)(U(r, X_r)\nabla X_r + \nabla z^{(\varepsilon)}_r) \qquad m \times \mathbb{P} - a.e; \end{aligned}$$

where $U(r, x)$ is defined in (2.7); see the equality just above equation (4.19) in [MZ02]. Integrating with respect to v over $v \in [t_i, t_j)$, on the one hand, and between $v \in [t_i, r)$, on the other, which yields

$$\frac{1}{t_j - t_i}\int_{t_i}^{t_j} D_v f^{(\varepsilon)}(\Theta_r)\sigma^{-1}(v, X_v)\nabla X_v dv = \frac{1}{r - t_i}\int_{t_i}^{r} D_v f^{(\varepsilon)}(\Theta_r)\sigma^{-1}(v, X_v)\nabla X_v dv.$$

One then follows the proof of [MZ02, Theorem 4.2], which essentially uses integration-by-parts for Malliavin calculus – Lemma 2.2 – to show that $\mathbb{E}_i[f^{(\varepsilon)}(r, X_r, Y^{(\varepsilon)}_r, Z^{(\varepsilon)}_r)H^{t_i}_r] = \mathbb{E}_i[f^{(\varepsilon)}(r, X_r, Y^{(\varepsilon)}_r, Z^{(\varepsilon)}_r)H^{t_i}_{t_j}]$. One extends to the general case (5.3) by convergence arguments as in the proof of Theorem 2.16 The relation (5.4) is now straightforward to obtain from (5.3). ☐

Lemma 5.3. *There is a finite constant C depending only on the bound on b and its derivatives, the bound on σ and its derivatives, L_f, θ_L, C_f, θ_c, and T such that, for all $i \in \{0, \ldots, N-1\}$,*

$$\|\mathbb{E}_{t_i}[\int_{t_i}^{t_{i+1}} f(r, X_r, Y_r, Z_r) H_r^{t_i} dr]\|_2 \leq C \int_{t_i}^{t_{i+1}} \frac{dr}{(T-r)^{1-\gamma}\sqrt{r-t_i}}, \tag{5.5}$$

$$\begin{aligned}&\|\sum_{j=i+1}^{N-1} \mathbb{E}_{t_i}[(f_j(Y_{t_{j+1}}, Z_{t_j}) - f_j(\bar{Y}_{j+1}, \bar{Z}_j)) H_{t_j}^{t_i}]\Delta_j\|_2 \\ &\qquad \leq C \sum_{j=i+1}^{N-1} \frac{\|Y_{t_{j+1}} - \bar{Y}_{j+1}\|_2 + \|Z_{t_j} - \bar{Z}_j\|_2}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j - t_i}} \Delta_j,\end{aligned} \tag{5.6}$$

$$\begin{aligned}&\|\sum_{j=i+1}^{N-1} \mathbb{E}_{t_i}[\int_{t_j}^{t_{j+1}} (f(r, X_r, Y_r, Z_r) - f(t_j, X_{t_j}, Y_{t_{j+1}}, Z_{t_j})) H_{t_j}^{t_i} dr]\|_2 \\ &\qquad \leq \frac{CN^{-1/2}}{(T-t_i)^{(1-\theta_X)/2}} + C \sum_{j=i+1}^{N-1} \frac{\int_{t_j}^{t_{j+1}} \{\|Y_r - Y_{t_{j+1}}\|_2 + \|Z_r - Z_{t_j}\|_2\} dr}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j - t_i}},\end{aligned} \tag{5.7}$$

where $\gamma := (\theta_c \wedge \frac{\alpha}{2} + \frac{\theta_L}{2}) \wedge \theta_c$.

Proof. In what follows, C may change from line to line. Using the conditional Cauchy-Schwarz inequality (Corollary 2.11),

$$\mathbb{E}_i[\int_{t_i}^{t_{i+1}} f(r, X_r, Y_r, Z_r) H_r^{t_i}] dr \leq \sqrt{C_M} \int_{t_i}^{t_{i+1}} \frac{(\mathbb{E}_i[|f(r, X_r, Y_r, Z_r)|^2])^{1/2}}{\sqrt{r-t_i}} dr;$$

then, Minkowski's inequality and the moment bound (2.22) of Corollary 2.13 imply that

$$\|\mathbb{E}_{t_i}[\int_{t_i}^{t_{i+1}} f(r, X_r, Y_r, Z_r) H_r^{t_i} dr]\|_2 \leq \sqrt{C_M} \int_{t_i}^{t_{i+1}} \frac{\|f(r, X_r, Y_r, Z_r)\|_2}{\sqrt{r-t_i}} dr \leq C \int_{t_i}^{t_{i+1}} \frac{dr}{(T-r)^{1-\gamma}\sqrt{r-t_i}}.$$

Using the Lipschitz continuity of f, Minkowski's inequality, and Lemma 2.10,

$$\|\sum_{j=i+1}^{N-1} \mathbb{E}_{t_i}[\big(f_j(Y_{t_{j+1}}, Z_{t_j}) - f_j(\bar{Y}_{j+1}, \bar{Z}_j)\big) H_{t_j}^{t_i}]\Delta_{j-1}\|_2 \leq C \sum_{j=i+1}^{N-1} \frac{\|Y_{t_{j+1}} - \bar{Y}_{j+1}\|_2 + \|Z_{t_j} - \bar{Z}_j\|_2}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j - t_i}} \Delta_j$$

For (5.7), the t-Hölder continuity of f in $(\mathbf{A_{f_t}})$, the Cauchy-Schwarz inequality (Corollary 2.11), Minkowski's inequality, and Hölder's inequality are needed:

$$\begin{aligned}&\|\sum_{j=i+1}^{N-1} \mathbb{E}_{t_i}[\int_{t_j}^{t_{j+1}} \big(f(r, X_r, Y_r, Z_r) - f_j(X_{t_j}, Y_{t_{j+1}}, Z_{t_j})\big) H_{t_j}^{t_i} dr]\|_2 \\ &\leq C \sum_{j=i+1}^{N-1} \frac{\int_{t_j}^{t_{j+1}} \|f(r, X_r, Y_r, Z_r) - f_j(X_r, Y_r, Z_r)\|_2 dr}{\sqrt{t_j - t_i}} + C \sum_{j=i+1}^{N-1} \frac{\int_{t_j}^{t_{j+1}} \|f_j(X_r, Y_r, Z_r) - f_j(X_{t_j}, Y_{t_{j+1}}, Z_{t_j})\|_2 dr}{\sqrt{t_j - t_i}} \\ &\leq C \sum_{j=i+1}^{N-1} \frac{\int_{t_j}^{t_{j+1}} \sqrt{r-t_j} dr}{\sqrt{t_j - t_i}} + C \sum_{j=i+1}^{N-1} \frac{\int_{t_j}^{t_{j+1}} \|X_r - X_{t_j}\|_2 dr}{(T-t_j)^{1-\theta_X/2}\sqrt{t_j - t_i}} + C \sum_{j=i+1}^{N-1} \frac{\int_{t_j}^{t_{j+1}} \{\|Y_r - Y_{t_{j+1}}\|_2 + \|Z_r - Z_{t_j}\|_2\} dr}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j - t_i}}\end{aligned}$$

The usual upper bound $\|X_r - X_{t_j}\|_2 \leq C\sqrt{r-t_j}$ implies that

$$\int_{t_j}^{t_{j+1}} \|X_r - X_{t_j}\|_2 dr \leq C \int_{t_j}^{t_{j+1}} \sqrt{r-t_j} dr.$$

Now, we obtain the upper bound $\int_{t_j}^{t_{j+1}} \sqrt{r-t_j}dr = \frac{2}{3}\Delta_j^{3/2} \leq CN^{-1/2}\Delta_j$ from Lemma B.1, and substitute it to the already acquired estimates to obtain

$$\begin{aligned}
&\| \sum_{j=i+1}^{N-1} \mathbb{E}_{t_i}[\int_{t_j}^{t_{j+1}} \big(f(r,X_r,Y_r,Z_r) - f_j(X_{t_j},Y_{t_{j+1}},Z_{t_j})\big) H_{t_j}^{t_i} dr]\|_2 \\
&\leq CN^{-1}\sum_{j=i+1}^{N-1} \frac{\Delta_j}{\sqrt{t_j-t_i}} + CN^{-1}\sum_{j=i+1}^{N-1} \frac{\Delta_j}{(T-t_j)^{1-\theta_X/2}\sqrt{t_j-t_i}} + C\sum_{j=i+1}^{N-1} \frac{\int_{t_j}^{t_{j+1}}\{\|Y_r - Y_{t_{j+1}}\|_2 + \|Z_r - Z_{t_j}\|_2\}dr}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j-t_i}}
\end{aligned}$$

Applying Lemma C.2 to bound the sums without the integrals is then sufficient to complete the proof. □

In the following proposition, we obtain a bound for the error terms on the right hand side of (5.7); these error terms are intrinsically related to the discritization error of the Malliavin weights scheme. Proposition 4.2 will be essential in the proof of this result.

Proposition 5.4. *Recall the definition $\gamma := (\theta_c \wedge \frac{\alpha}{2} + \frac{\theta_L}{2}) \wedge \theta_c$. Let either $(\mathbf{A_{exp\Phi}})$ or $(\mathbf{A_{h\Phi}})$ be in force and suppose that $0 < \beta < (2\gamma) \wedge \alpha \wedge \theta_L$. For $\delta, K > 0$, define $\mathcal{C}(\delta,K) := K(N^{-1/2}\mathbf{1}_{[1,3]}(\delta) + N^{-\gamma}\mathbf{1}_{(0,1)}(\delta))$, and, for $j \in \{0,\ldots,N-1\}$,*

$$\Psi_j := \int_{t_j}^{t_{j+1}} \{\|Y_r - Y_{t_{j+1}}\|_2 + \|Z_r - Z_{t_j}\|_2\}dr.$$

There is a constant C depending only on L_f, θ_L, C_f, θ_c, β, $\bar{\beta}$, the bound on b and its derivatives, the bound on σ and it's derivatives, and T, but not on N, such that, for all $N \geq 1$,

$$\begin{aligned}
\sum_{j=0}^{N-1} \frac{\Psi_j}{(T-t_j)^{(1-\theta_L)/2}} &\leq C(T-t_i)^{(1+\theta_L-\beta)/2}\mathcal{C}(\beta+2\gamma, \ln(N)^{1/4} \vee 1), \\
\sum_{j=i+1}^{N-1} \frac{\Psi_j}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j-t_i}} &\leq C(T-t_i)^{-(1+\beta-\theta_L)/2}\mathcal{C}(\beta+2\gamma, \ln(N)^{1/4} \vee 1)
\end{aligned}$$

in the case of $(\mathbf{A_{exp\Phi}})$, and

$$\begin{aligned}
\sum_{j=0}^{N-1} \frac{\Psi_j}{(T-t_j)^{(1-\theta_L)/2}} &\leq C(T-t_i)^{(1+\theta_L-\beta)/2}\mathcal{C}(\beta+\theta_\Phi+2\gamma, K_\Phi), \\
\sum_{j=i+1}^{N-1} \frac{\Psi_j}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j-t_i}} &\leq C(T-t_i)^{-(1+\beta-\theta_L)/2}\mathcal{C}(\beta+\theta_\Phi+2\gamma, K_\Phi)
\end{aligned}$$

in the case of $(\mathbf{A_{h\Phi}})$.

Proof. We will prove the bounds for

$$\sum_{j=i+1}^{N-1} \frac{\Psi_j}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j-t_i}}.$$

The bounds for the $\sum_{j=0}^{N-1} \frac{\Psi_j}{(T-t_j)^{(1-\theta_L)/2}}$ are obtained analogously. Moreover, we will only prove the result for the terms in Z. The bound for the terms in Y are also obtained analogously. In what follows, C may change from line to line. We first prove the result under $(\mathbf{A_{\partial f}})$ and $(\mathbf{A_{b\Phi}})$, and then obtain the general result by means of mollification. Fix $\varepsilon \leq \Delta_{N-1}$ and recall the BSDE

$(Y^{(\varepsilon)}, Z^{(\varepsilon)})$ from Definition 2.7 in Section 2.3. We use the version of $Z^{(\varepsilon)}$ provided by Theorem 2.16. First, apply the triangle inequality to the integrand in order to obtain $\|Z_t - Z_{t_i}\|_2 \le \|Z_t - Z_t^{(\varepsilon)}\|_2 + \|Z_{t_i} - Z_{t_i}^{(\varepsilon)}\|_2 + \|Z_t^{(\varepsilon)} - Z_{t_i}^{(\varepsilon)}\|_2$.

To bound the terms in $Z - Z^{(\varepsilon)}$, recall the bound (2.24) from Corollary 2.14. For $j \le N-2$, the bound on $\|Z_t - Z_t^{(\varepsilon)}\|_2$ implies that

$$\int_{t_j}^{t_{j+1}} \|Z_t - Z_t^{(\varepsilon)}\|_2 dt \le C \int_{t_j}^{t_{j+1}} \frac{\int_{T-\varepsilon}^{T} (T-r)^{\gamma-1} dr}{\sqrt{t_{N-1} - t}} dt \le C\varepsilon^{\gamma} \int_{t_j}^{t_{j+1}} \frac{dt}{\sqrt{t_{N-1} - t}}.$$

Lemma C.1 yields $\int_{t_j}^{t_{j+1}} (t_{N-1} - t)^{-1/2} dt \le 2\Delta_j (t_{N-1} - t_j)^{-1/2}$. Therefore, applying Lemma C.2 implies that

$$\sum_{j=i+1}^{N-2} \frac{\int_{t_j}^{t_{j+1}} \|Z_t - Z_t^{(\varepsilon)}\|_2 dt}{(T-t_j)^{(1-\theta_L)/2} \sqrt{t_j - t_i}} \le C\varepsilon^{\gamma} \sum_{j=i+1}^{N-2} \frac{\Delta_j}{(t_{N-1} - t_j)^{1-\theta_L/2} \sqrt{t_j - t_i}} \le \frac{C\varepsilon^{\gamma}}{(t_{N-1} - t_i)^{(1-\theta_L)/2}}. \tag{5.8}$$

Then, use $(t_{N-1} - t_i)^{-1/2} \le 2(T - t_i)^{-1/2}$ on the denominator on the right hand side. For the outstanding term, $j = N-1$, we implement Lemma C.2 to show that

$$\int_{t_{N-1}}^{T} \|Z_t - Z_t^{(\varepsilon)}\|_2 dt \le C \int_{t_{N-1}}^{T} \Big\{ \int_t^T (T-r)^{\gamma-1} (r-t)^{-1/2} dr \Big\} dt \le C \Delta_{N-1}^{1/2+\gamma},$$

whence it follows that

$$\frac{\int_{t_{N-1}}^{T} \|Z_t - Z_t^{(\varepsilon)}\|_2 dt}{\Delta_{N-1}^{(1-\theta_L)/2} \sqrt{t_{N-1} - t_i}} \le \frac{C\Delta_{N-1}^{\gamma+\theta_L/2}}{\sqrt{t_{N-1} - t_i}} \le \frac{C\Delta_{N-1}^{\gamma+\theta_L/2}}{\sqrt{T - t_i}} \tag{5.9}$$

Combining (5.8) and (5.9), it follows that

$$\sum_{j=i+1}^{N-1} \frac{\int_{t_j}^{t_{j+1}} \|Z_t - Z_t^{(\varepsilon)}\|_2 dr}{(T-t_j)^{(1-\theta_L)/2} \sqrt{t_j - t_i}} \le \frac{C\varepsilon^{\gamma}}{(T-t_i)^{(1-\theta_L)/2}} + \frac{CN^{-1}}{\sqrt{T - t_i}} \tag{5.10}$$

where we have used that $\Delta_{N-1}^{2\gamma+\theta_L} = TN^{(2\gamma+\theta_L)/\beta}$ and $\beta < (2\gamma) \wedge \theta_L$. Analogously, we can also show that

$$\sum_{j=i+1}^{N-1} \frac{\|Z_{t_j} - Z_{t_j}^{(\varepsilon)}\|_2 \Delta_j}{(T-t_j)^{(1-\theta_L)/2} \sqrt{t_j - t_i}} \le \frac{C\varepsilon^{\gamma}}{(T-t_i)^{(1-\theta_L)/2}} + \frac{CN^{-1}}{\sqrt{T - t_i}} \tag{5.11}$$

Recalling the BSDEs (y, z) and $(y^{(\varepsilon)}, z^{(\varepsilon)})$ from Definition 2.7 and that $Z^{(\varepsilon)} = z + z^{(\varepsilon)}$, the triangle inequality yields $\|Z_t^{(\varepsilon)} - Z_{t_i}^{(\varepsilon)}\|_2 \le \|z_t - z_{t_i}\|_2 + \|z_t^{(\varepsilon)} - z_{t_i}^{(\varepsilon)}\|_2$. In the proof of [GM10, Theorem 1.1], in bounding the terms E_1 and E_2, it is shown that, for all $t \in [0, T]$,

$$\|z_t - z_{t_i}\|_2^2 \le \frac{C(t - t_i)}{(T-t)^{1-\alpha}} + C \int_{t_i}^{t} \|\nabla_x^2 u(r, X_r)\|_2^2 dr.$$

Lemma 2.8 implies $\int_{t_i}^{t} \|\nabla_x^2 u(r, X_r)\|_2^2 dr \le C \int_{t_i}^{t} (T-r)^{\alpha-2} dr$. Now, applying Jensen's inequality, Lemma C.1, Lemma C.2, and the above bound, one obtains

$$\begin{aligned}
\sum_{j=i+1}^{N-1} \frac{\int_{t_j}^{t_{j+1}} \|z_r - z_{t_j}\|_2 dr}{(T-t_j)^{(1-\theta_L)/2} \sqrt{t_j - t_i}} &\le \sum_{j=i+1}^{N-1} \frac{C\Delta_j \int_{t_j}^{t_{j+1}} (T-r)^{(\alpha-1)/2} dr + C \int_{t_j}^{t_{j+1}} \Big(\int_{t_j}^{r} (T-t)^{\alpha-2} dt \Big)^{1/2} dr}{(T-t_j)^{(1-\theta_L)/2} \sqrt{t_j - t_i}} \\
&\le \sum_{j=i+1}^{N-1} \frac{C\Delta_j^2 (T-t_j)^{(\alpha-1)/2} + C\Delta_j^{1/2} \Big(\int_{t_j}^{t_{j+1}} (t_{j+1} - t)(T-t)^{\alpha-2} dt \Big)^{1/2}}{(T-t_j)^{(1-\theta_L)/2} \sqrt{t_j - t_i}}.
\end{aligned} \tag{5.12}$$

For $j \le N-2$, one can apply Lemma B.1 and Lemma C.1 to show that

$$\int_{t_j}^{t_{j+1}} (t_{j+1}-t)(T-t)^{\alpha-2}dt \le \frac{\Delta_j}{(T-t_{j+1})^{1-\beta}} \int_{t_j}^{t_{j+1}} (T-t)^{\alpha-\beta-1}dt \le \frac{CN^{-1}\Delta_j}{(T-t_j)^{1-\alpha+\beta}}.$$

On the other hand, for $j = N-1$, since $\beta < \alpha$,

$$\int_{t_{N-1}}^{T} (T-t)(T-t)^{\alpha-2}dt = \frac{1}{\alpha}\Delta_{N-1}^{\alpha} = \frac{T}{\alpha}N^{-\alpha/\beta} \le \frac{T}{\alpha}N^{-1}.$$

Substituting these bounds into (5.12) and implementing Lemma B.1 and Lemma C.2, we obtain

$$\begin{aligned}\sum_{j=i+1}^{N-1} \frac{\int_{t_j}^{t_{j+1}} \|z_r - z_{t_j}\|_2 dr}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j-t_i}} &\le CN^{-1/2}\sum_{j=i+1}^{N-1} \frac{\Delta_j}{(T-t_j)^{1+(\beta-\alpha-\theta_L)/2}\sqrt{t_j-t_i}} + CN^{-1/2}\frac{\Delta_{N-1}^{\theta_L/2}}{\sqrt{t_{N-1}-t_i}} \\ &\le \frac{CN^{-1/2}}{(T-t_i)^{(1+\beta-\alpha-\theta_L)/2}} + \frac{CN^{-1}}{\sqrt{T-t_i}}. \end{aligned} \tag{5.13}$$

In the bounds (3.2), we used the inequality

$$\|z_r^{(\varepsilon)} - z_{t_i}^{(\varepsilon)}\|_2 \le C\int_{t_i}^{r} \|a_t^{(\varepsilon)}\|_2 dt + C\int_{t_i}^{r} \|V_t^{(\varepsilon)}\|_2 dt + C\Delta_i^{1/2}. \tag{5.14}$$

Using $\|a_t^{(\varepsilon)}\|_2 \le C(T-t)^{(\alpha+\theta_L-3)/2}$ as shown Lemma 2.9, Lemma B.1, Lemma C.1, and Lemma C.2, it follows that

$$\begin{aligned}&\sum_{j=i+1}^{N-1} \frac{\int_{t_j}^{t_{j+1}}\{\int_{t_j}^{r} \|a_t^{(\varepsilon)}\|_2 dt\}dr}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j-t_i}} \le C\sum_{j=i+1}^{N-1} \frac{\int_{t_j}^{t_{j+1}}\{\int_{t_j}^{r} (T-t)^{(\theta_L+\alpha-3)/2}dt\}dr}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j-t_i}} \\ &\le C\sum_{j=i+1}^{N-1} \frac{\int_{t_j}^{t_{j+1}}(T-r)^{(\alpha-2)/2}dr \int_{t_i}^{t_{j+1}}(T-t)^{(\theta_L-1)/2}dt}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j-t_i}} \le C\sum_{j=i+1}^{N-1} \frac{\Delta_j^{(3+\theta_L)/2}}{(T-t_j)^{(3-\alpha-\theta_L)/2}\sqrt{t_j-t_i}} \\ &\le C\Big(\max_{0\le i\le N-1} \frac{\Delta_i^{1+\theta_L}}{(T-t_i)^{1-\beta}}\Big)^{1/2} \sum_{j=i+1}^{N-1} \frac{\Delta_j}{(T-t_j)^{(2+\beta-\theta_L-\alpha)/2}\sqrt{t_j-t_i}} \le \frac{CN^{-1/2}}{(T-t_i)^{(1+\beta-\theta_L-\alpha)/2}} \end{aligned} \tag{5.15}$$

On the other hand, we obtain bounds for $\|V_t^{(\varepsilon)}\|_2$ from Proposition 4.2 under $(\mathbf{A_{exp\Phi}})$ or $(\mathbf{A_{h\Phi}})$. Let us work under $(\mathbf{A_{exp\Phi}})$. It follows from Lemma C.1 and Lemma C.2 that, for all j and $r \in [t_j, t_{j+1}]$,

$$\begin{aligned}\int_{t_j}^{r} \|V_t^{(\varepsilon)}\|_2 dt &\le C\|\Phi\|_\infty \int_{t_j}^{r} \Big\{\int_{t}^{T-\varepsilon} (T-s)^{(\beta-1)/2}(T-s)^{(\theta_L-\beta-2)/2}(s-t)^{-1/2}ds\Big\}dt \\ &\le C\|\Phi\|_\infty \varepsilon^{(\beta-1)/2}\int_{t_j}^{r} \Big\{\int_{t}^{T} (T-s)^{\theta_L-\beta-1}(s-t)^{-1/2}ds\Big\}dt \\ &\le C\|\Phi\|_\infty \varepsilon^{(\beta-1)/2}\int_{t_j}^{r} (T-t)^{(\theta_L-\beta-1)/2}dt \le C\|\Phi\|_\infty \frac{\varepsilon^{(\beta-1)/2}\Delta_j}{(T-t_j)^{(1+\beta-\theta_L)/2}}\end{aligned}$$

Therefore, using Lemma B.1, Lemma C.2 and the above bound,

$$\begin{aligned}\sum_{j=i+1}^{N-1} \frac{\int_{t_j}^{t_{j+1}}\{\int_{t_j}^{r} \|V_t^{(\varepsilon)}\|_2 dt\}dr}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j-t_i}} &\le C\|\Phi\|_\infty \varepsilon^{(\beta-1)/2}\max_i \Delta_i \sum_{j=i+1}^{N-1} \frac{\Delta_j}{(T-t_j)^{1-(\theta_L-\beta)/2}\sqrt{t_j-t_i}} \\ &\le \frac{C\|\Phi\|_\infty \varepsilon^{(\beta-1)/2}N^{-1}}{(T-t_i)^{(1+\beta-\theta_L)/2}}. \end{aligned} \tag{5.16}$$

Now, substituting (5.15) and (5.16) into (5.14), it follows that

$$\sum_{j=i+1}^{N-1} \frac{\int_{t_j}^{t_{j+1}} \|z_r^{(\varepsilon)} - z_{t_j}^{(\varepsilon)}\|_2 dr}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j - t_i}} \leq \frac{CN^{-1/2}}{(T-t_i)^{(1+\beta-\theta_L-\alpha)/2}} + \frac{C\|\Phi\|_\infty \varepsilon^{(\beta-1)/2} N^{-1}}{(T-t_i)^{(1+\beta-\theta_L)/2}} \tag{5.17}$$

Combining (5.10), (5.11), (5.13) and (5.17) yields

$$\sum_{j=i+1}^{N-1} \frac{\int_{t_j}^{t_{j+1}} \|Z_r - Z_{t_j}\|_2 dr}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j - t_i}}$$
$$\leq \frac{CN^{-1/2}}{(T-t_i)^{(1+\beta-\theta_L-\alpha)/2}} + \frac{CN^{-1}}{\sqrt{T-t_i}} + \frac{C\varepsilon^\gamma}{(T-t_i)^{(1-\theta_L)/2}} + \frac{C\|\Phi\|_\infty \varepsilon^{(\beta-1)/2} N^{-1}}{(T-t_i)^{(1+\beta-\theta_L)/2}}$$

and we take $\varepsilon = N^{-1/(2\gamma)}$ if $1-\beta-2\gamma<0$ and $\varepsilon = N^{-1}$ otherwise to complete the proof under ($\mathbf{A_{\partial f}}$) and ($\mathbf{A_{b\Phi}}$). The proof under ($\mathbf{A_{\partial f}}$) and ($\mathbf{A_{h\Phi}}$)is analogous.

To prove the result without ($\mathbf{A_{\partial f}}$) or ($\mathbf{A_{b\Phi}}$), recall the mollified BSDE (Y_M, Z_M) from Corollary 2.15. Set $M = (3\ln(N))^{1/4}$ and $R(M)$ equal to $3L_f e^{M^2/2}$. Substituting equations (4.15) and (4.16) into (4.14), $\|Z_s - Z_{M,s}\|_2 \leq CN^{-1}(T-s)^{-1/2}$ for all $s \in [0,T)$, whence the triangle inequality and Lemma C.1 imply

$$\int_{t_j}^{t_{j+1}} \|Z_r - Z_{t_j}\|_2 dr \leq \int_{t_j}^{t_{j+1}} \|Z_r - Z_{M,r}\|_2 dr + \|Z_{t_j} - Z_{M,t_j}\|_2 \Delta_j + \int_{t_j}^{t_{j+1}} \|Z_{M,r} - Z_{M,t_j}\|_2 dr$$
$$\leq CN^{-1}\Delta_j (T-t_j)^{-1/2} + \int_{t_j}^{t_{j+1}} \|Z_{M,r} - Z_{M,t_j}\|_2 dr.$$

The proof is then completed with use of Lemma C.2. □

We come to the main result of this section, namely the error estimation for the Malliavin weights scheme.

Theorem 5.5. *Recall the definition* $\gamma := (\theta_c \wedge \frac{\alpha}{2} + \frac{\theta_L}{2}) \wedge \theta_c$. *Let* ($\mathbf{A_{exp\Phi}}$) *or* ($\mathbf{A_{h\Phi}}$) *be and force and suppose that* $0 < \beta < \gamma \wedge \alpha \wedge \theta_L$. *For* $\delta, K > 0$, *define* $\mathcal{C}(\delta, K) := KN^{-1/2}\mathbf{1}_{[1,3]}(\delta) + N^{-\gamma}\mathbf{1}_{(0,1)}(\delta)$. *There is a constant* C *depending only on* L_f, θ_L, C_f, θ_c, β, $\bar{\beta}$, *the bound on* b *and its derivatives, the bound on* σ *and it's derivatives,* $K^\alpha(\Phi)$ *and* T, *but not on* N, *such that, for all* $N \geq 1$,

$$\left.\begin{array}{l} \|Y_{t_i} - \bar{Y}_i^{(N)}\|_2 \leq C(T-t_i)^{(1+\theta_L-\beta)/2}\mathcal{C}(\beta+2\gamma, \ln(N)^{1/4} \vee 1), \\ \|Z_{t_i} - \bar{Z}_i^{(N)}\|_2 \leq C(T-t_i)^{-(1+\beta-\theta_L)/2}\mathcal{C}(\beta+2\gamma, \ln(N)^{1/4} \vee 1) \\ \qquad +CN^{-1/2}(T-t_i)^{-(1-\alpha\wedge(2\gamma)\wedge\theta_X)/2}. \end{array}\right\} \text{ in the case of } (\mathbf{A_{exp\Phi}}),$$

$$\left.\begin{array}{l} \|Y_{t_i} - \bar{Y}_i^{(N)}\|_2 \leq C(T-t_i)^{(1+\theta_L-\beta)/2}\mathcal{C}(\beta+\theta_\Phi+2\gamma, K_\Phi), \\ \|Z_{t_i} - \bar{Z}_i^{(N)}\|_2 \leq C(T-t_i)^{-(1+\beta-\theta_L)/2}\mathcal{C}(\beta+\theta_\Phi+2\gamma, K_\Phi) \\ \qquad +CN^{-1/2}(T-t_i)^{-(1-\alpha\wedge(2\gamma)\wedge\theta_X)/2} \end{array}\right\} \text{ in the case of } (\mathbf{A_{h\Phi}}).$$

Proof. In what follows, C may change from line to line. For simplicity, we omit the process X from the driver, so that $f(t,y,z) := f(t,X_t,y,z)$ and $f_j(y,z) := f_j(X_{t_j},y,z)$. Fix $i \in \{0,\ldots,N-1\}$. Using the estimates from Lemma 5.1 and Lemma 5.3, and (5.4) from Lemma 5.2, it follows that

$$\|Z_{t_i} - \bar{Z}_i^{(N)}\|_2 = \|\mathbb{E}_i[\Phi(X_T)H_T^{t_i} - \Phi(X_T)H_N^i + \int_{t_i}^T f(t, Y_t, Z_t)H_t^{t_i}dt - \sum_{j=i+1}^{N-1} f_j(\bar{Y}_{j+1}^{(N)}, \bar{Z}_j^{(N)})H_j^i\Delta_j]\|_2$$

$$\le \|\mathbb{E}_i[\int_{t_i}^{t_{i+1}} f(r, Y_r, Z_r)H_r^{t_i}dr]\|_2 + \|\mathbb{E}_i[\Phi(X_T)(H_T^{t_i} - H_N^i)]\|_2 + \|\sum_{j=i+1}^{N-1}\mathbb{E}_i[f_j(Y_{t_{j+1}}, Z_{t_j})(H_{t_j}^{t_i} - H_j^i)]\Delta_j\|_2$$

$$+ \|\sum_{j=i+1}^{N-1}\mathbb{E}_i[(f_j(Y_{t_{j+1}}, Z_{t_j}) - f_j(\bar{Y}_{j+1}^{(N)}, \bar{Z}_j^{(N)}))H_j^i]\Delta_j\|_2 + \|\sum_{j=i+1}^{N-1}\mathbb{E}_i[\int_{t_j}^{t_{j+1}}(f(r, Y_r, Z_r) - f_j(Y_{t_{j+1}}, Z_{t_j}))H_{t_j}^{t_i}dr]\|_2$$

$$\le \frac{CN^{-1/2}}{(T - t_i)^{(1-\alpha\wedge(2\gamma)\wedge\theta_X)/2}} + C\int_{t_i}^{t_{i+1}}\frac{dr}{(T-r)^{1-\gamma}\sqrt{r - t_i}} + C\mathcal{H}(i) + C\sum_{j=i+1}^{N-1}\frac{\Theta_j\Delta_j}{(T - t_j)^{(1-\theta_L)/2}\sqrt{t_j - t_i}} \tag{5.18}$$

where $\Theta_j := \|Y_{t_{j+1}} - \bar{Y}_{j+1}^{(N)}\|_2 + \|Z_{t_j} - \bar{Z}_j^{(N)}\|_2$ and

$$\mathcal{H}(i) := \sum_{j=i+1}^{N-1}\frac{\Psi(j)}{(T - t_j)^{(1-\theta_L)/2}\sqrt{t_j - t_i}}, \quad \Psi(j) := \int_{t_j}^{t_{j+1}}\{\|Y_r - Y_{t_{j+1}}\|_2 + \|Z_r - Z_{t_j}\|_2\}dr.$$

In (5.18), we have estimated $\|\sum_{j=i+1}^{N-1}\mathbb{E}_i[f_j(Y_{t_{j+1}}, Z_{t_j})(H_{t_j}^{t_i} - H_j^i)]\Delta_j\|_2$ by $CN^{-1/2}\sum_{j=i+1}^{N-1}(T - t_j)^{\gamma-1}(T_j - t_i)^{-1/2}\Delta_j$ using Lemma 5.1, and the latter sum by $C(T - t_i)^{\gamma-1/2}$ using Lemma C.2. Using a similar technique, $\|Y_{t_i} - \bar{Y}_i^{(N)}\|_2$ is bounded above by

$$\begin{aligned} & C\int_{t_i}^{t_{i+1}}\frac{dr}{(T-r)^{1-\gamma}} + C\sum_{j=i}^{N-1}\frac{\Psi(j)}{(T-t_j)^{(1-\theta_L)/2}} + C\sum_{j=i}^{N-1}\frac{\Theta_j\Delta_j}{(T-t_j)^{(1-\theta_L)/2}} \\ \le \quad & CN^{-1/2} + C\sum_{j=i}^{N-1}\frac{\Psi(j)}{(T-t_j)^{(1-\theta_L)/2}} + C\sum_{j=i}^{N-1}\frac{\Theta_j\Delta_j}{(T-t_j)^{(1-\theta_L)/2}} \end{aligned} \tag{5.19}$$

where we have used Lemma B.1 and Lemma C.1 on the first integral to obtian $\int_{t_i}^{t_{i+1}}(T-r)^{\gamma-1}dr \le C\Delta_j(T - t_j)^{\gamma-1} \le CN^{-1/2}$. It follows from (5.18) and (5.19) that

$$\Theta_i \le \frac{CN^{-1/2}}{(T - t_i)^{(1-\alpha\wedge(2\gamma)\wedge\theta_X)/2}} + C\int_{t_i}^{t_{i+1}}\frac{dr}{(T-r)^{1-\gamma}\sqrt{r - t_i}} + C\mathcal{H}(i) + C\sum_{j=i+1}^{N-1}\frac{\Theta_j\Delta_j}{(T - t_j)^{(1-\theta_L)/2}\sqrt{t_j - t_i}}$$

Letting $U_i := \Theta_i$, $\Gamma(i) := N^{-1/2}(T - t_i)^{(\alpha\wedge(2\gamma)\wedge\theta_X-1)/2}$, $\Xi(i) := \int_{t_i}^{t_{i+1}}\frac{dr}{(T-r)^{1-\gamma}\sqrt{r-t_i}}$, and

$$W_i := \Gamma(i) + \Xi(i) + \mathcal{H}(i), \tag{5.20}$$

it follows from Lemma C.3 that

$$\Theta_i \le CW_i + C\sum_{j=i+1}^{N-1}\frac{W_j\Delta_j}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j - t_i}} + C\sum_{j=i+1}^{N-1}\frac{\Theta_j\Delta_j}{(T-t_j)^{(1-\theta_L)/2}} \tag{5.21}$$

Therefore, using Lemma C.4 in (5.18) and (5.19),

$$\|Z_{t_i} - \bar{Z}_i^{(N)}\|_2 \le CW_i + C\sum_{j=i+1}^{N-1}\frac{W_j\Delta_j}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j - t_i}}, \tag{5.22}$$

$$\|Y_{t_i} - \bar{Y}_i^{(N)}\|_2 \le CN^{-1/2} + C\sum_{j=i}^{N-1}\frac{\Psi(j)}{(T-t_j)^{(1-\theta_L)/2}} + C\sum_{j=i}^{N-1}\frac{W_j\Delta_j}{(T-t_j)^{(1-\theta_L)/2}}. \tag{5.23}$$

Let us consider the sum in the W terms. Firstly, remark that we only need consider the sums for $i < N-1$. Recall the terminology of (5.20). Using Lemma C.2,

$$\sum_{j=i+1}^{N-1} \frac{\Gamma(j)\Delta_j}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j - t_i}} = CN^{-1/2} \sum_{j=i+1}^{N-1} \frac{\Delta_j}{(T-t_j)^{1-\theta_L/2}\sqrt{t_j-t_i}} \le CN^{-1/2}(T-t_i)^{(\theta_L-1)/2}. \tag{5.24}$$

Using the fact that $\Delta_j \le \Delta_{j-1}$ to show that $\sqrt{t_{j+1}-t_i}/\sqrt{t_j-t_i} \le 2$, Lemma B.1 to show that $\Delta_j/\Delta_{j+1} \le C$ and $\max_j \Delta_j(T-t_j)^{2\gamma-1} \le N^{-1}$, one can apply Lemma C.2 to bound the sum in $\Xi(j)$ as follows:

$$\begin{aligned}
&\sum_{j=i+1}^{N-1} \frac{\Xi(j)\Delta_j}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j-t_i}} \\
&\quad \le C\frac{\int_{t_{N-1}}^T (T-r)^{\gamma-1}(r-t_{N-1})^{-1/2}dr\Delta_{N-1}}{\Delta_{N-1}^{(1-\theta_L)/2}\sqrt{t_{N-1}-t_i}} + C\sum_{j=i+1}^{N-2} \frac{\Delta_{j+1}^{3/2}(\Delta_j/\Delta_{j+1})^{3/2}\sqrt{t_{j+1}-t_i}/\sqrt{t_j-t_i}}{(T-t_{j+1})^{(3-\theta_L-2\gamma)/2}\sqrt{t_{j+1}-t_i}} \\
&\quad \le \frac{C\Delta_{N-1}^{(1+\gamma+\theta_L)/2}\sqrt{\Delta_{N-2}}}{\sqrt{t_{N-1}-t_i}} + C\max_j \sqrt{\frac{\Delta_j}{(T-t_j)^{1-2\gamma}}} \sum_{j=i+1}^{N-2} \frac{\Delta_j}{(T-t_{j+1})^{1-\theta_L/2}\sqrt{t_j-t_i}} \\
&\quad \le CN^{-3/2} + CN^{-1/2}(T-t_i)^{(\theta_L-1)/2}
\end{aligned} \tag{5.25}$$

In order to deal with the sum in $\mathcal{H}(j)$, we change the order of summation and apply Lemma C.2 to obtain

$$\begin{aligned}
\sum_{j=i+1}^{N-1} \frac{\mathcal{H}(j)\Delta_j}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j-t_i}} &= \sum_{j=i+1}^{N-1} \frac{\sum_{k=j+1}^{N-1} \frac{\Psi(k)}{(T-t_k)^{(1-\theta_L)/2}\sqrt{t_k-t_j}}\Delta_j}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j-t_i}} \\
&= \sum_{k=i+2}^{N-1} \frac{\sum_{j=i+1}^{k-1} \frac{\Delta_j}{(T-t_j)^{1-\theta_L/2}\sqrt{t_j-t_i}}\Psi(k)}{(T-t_k)^{(1-\theta_L)/2}} \\
&\le C\sum_{j=i+1}^{N-1} \frac{\Psi(j)}{(T-t_j)^{(1-\theta_L)/2}(t_j-t_i)^{(1-\theta_L)/2}} = C\mathcal{H}(i).
\end{aligned} \tag{5.26}$$

Combining (5.24) - (5.26), the bound on the sum in W_j is

$$\sum_{j=i+1}^{N-1} \frac{W_j\Delta_j}{(T-t_j)^{(1-\theta_L)/2}\sqrt{t_j-t_i}} \le CN^{-1/2}(T-t_i)^{(\theta_L-1)/2} + C\mathcal{H}(i). \tag{5.27}$$

By analogous calculations, one shows that

$$\sum_{j=i+1}^{N-1} \frac{W_j\Delta_j}{(T-t_j)^{(1-\theta_L)/2}} \le CN^{-1/2}(T-t_i)^{(\theta_L+1)/2} + C\sum_{j=i}^{N-1} \frac{\Psi(j)}{(T-t_j)^{(1-\theta_L)/2}}. \tag{5.28}$$

The proof is completed by substituting (5.27) into (5.22), (5.28) into (5.23), and using Proposition 5.4 to bound the remaining terms. □

A Stochastic analysis

The following *conditional* Fubini's theorem is a consequence of the Monotone Class Theorem.

Lemma A.1. *Let $f_s \in \mathbf{L}_2([0,T]\times\Omega)$. Then, for all $t \in [0,T]$, there exists a $\mathcal{B}([0,T])\otimes\mathcal{F}_t$-measurable processes F_t belonging to $L_2([0,T]\times\Omega)$ such that $(\omega,s)\mapsto F_t(s)$ is a version of $(\omega,s)\mapsto \mathbb{E}_t[f_s]$ and*

$$\mathbb{E}_t[\int_0^T f_s ds] = \int_0^T F_t(\cdot,s)ds \quad \textit{almost surely.}$$

We need the following generalization of the a priori estimates [BDH+03, Proposition 3.2]:

Proposition A.2. *Let k be an integer, and p be an integer greater than or equal to 2. Let $f : \Omega\times[0,T)\times(\mathbb{R}^k)^\top\times\mathbb{R}^{q\times k} \to (\mathbb{R}^k)^\top$ be $\mathcal{P}\times\mathcal{B}\big((\mathbb{R}^k)^\top\big)\otimes\mathcal{B}(\mathbb{R}^{q\times k})$-measurable, and ξ be an $(\mathbb{R}^k)^\top$-valued random variable in $\mathbf{L}_p(\mathcal{F}_T)$. Let $(f_t)_{t\in[0,T]}$ be non-negative, predictable process, $\mu \in \mathbf{L}_1([0,T];m)$ and $\lambda \in \mathbf{L}_2([0,T];m)$ be $\mathbb{R}$-valued non-negative. Additionally, assume that $\mathbb{E}[(\int_0^T f_t dt)^p] < \infty$. For any $(y_1,y_2)\in(\mathbb{R}^k)^2$, define the scalar product $(y_1,y_2) := \sum_{j=1}^k y_{1,j}y_{2,j}$ and assume that, for all $(t,y,z)\in[0,T)\times(\mathbb{R}^k)^\top\times\mathbb{R}^{k\times q}$, $(\omega,t,y,z)\mapsto f(\omega,t,y,z)$ satisfies*

$$\big(|y|^{-1}y\mathbf{1}_{|y|>0}, f(\omega,t,y,z)\big) \le f_t(\omega) + \mu_t|y| + \lambda_t|z| \quad \textit{almost surely.} \tag{A.1}$$

Let (Y,Z) be a solution to the $\big((\mathbb{R}^k)^\top,\mathbb{R}^{q\times k}\big)$-valued BSDE

$$Y_t = \xi + \int_t^T f(r,Y_r,Z_r)dr - \sum_{j=1}^q \int_t^T (Z_{j,r})^\top dW_{j,r}.$$

in the space $\mathcal{S}^p\times\mathcal{H}^p$, where $\mathcal{H}^p$ is the space of predictable processes X such that $\mathbb{E}[(\int_0^T |X_s|^2 ds)^{p/2}]$ is finite; Z_j denotes the j-th column of Z.

Then, there exists a constant C_p, depending only on p, such that, for any $\eta_t \ge \mu_t + \lambda_t^2/(p-1)$ in $L_1(\mathbb{R};dt)$,

$$\mathbb{E}[\sup_t e^{p\int_0^t \eta_r dr}|Y_t|^p + (\int_0^T e^{2\int_0^t \eta_r dr}|Z_t|^2 dt)^{p/2}] \le C_p\mathbb{E}[e^{p\int_0^T \eta_r dr}|\xi|^p + (\int_0^T e^{\int_0^t \eta_r dr} f_t dt)^p].$$

Proof. Consider the processes $\tilde{Y}_t = e^{\int_0^t \eta_r dr}Y_t$ and $\tilde{Z}_t = e^{\int_0^t \eta_r dr}Z_t$. Then $(\tilde{Y},\tilde{Z})$ satisfies a BSDE with terminal condition $\tilde{\xi} = e^{\int_0^T \eta_r dr}\xi$ and driver $\tilde{f}(t,y,z) = e^{\int_0^t \eta_r dr} f(t, e^{-\int_0^t \eta_r dr}y, e^{-\int_0^t \eta_r dr}z) - \eta_t y$. Moreover, for all $(t,y,z)\in[0,T)\times\mathbb{R}^k\times\mathbb{R}^{k\times q}$, $\tilde{f}(\omega,y,z)$ satisfies

$$\big(|y|^{-1}y\mathbf{1}_{|y|>0}, \tilde{f}(\omega,t,y,z)\big) \le \tilde{f}_t(\omega) + \tilde{\mu}_t|y| + \tilde{\lambda}_t|z| \quad \text{almost surely.}$$

with $\tilde{f}_t = e^{-\int_0^t \eta_r dr} f_t$, $\tilde{\mu}_t = \mu_t - \eta_t$, and $\tilde{\lambda}_t = \lambda_t$. The rest of the proof follows exactly as the proof of [BDH+03, Proposition 3.2]. □

B Time-grids

Lemma B.1. *The time grid $\pi^{(\beta)} = \{0 = t_0 < \ldots < t_N = T \ : \ t_i = T - T(1-i/N)^{1/\beta}\}$ with $\beta \in (0,1]$ satisfies*

$$\max_{0\le i<N} \frac{\Delta_k}{(T-t_k)^{1-\theta}} \le \frac{T^\theta}{\beta}\frac{1}{N^{1\wedge\frac{\theta}{\beta}}}, \tag{B.1}$$

$$\max_{0\le i\le N-2} \frac{\Delta_k}{\Delta_{k+1}} \le \frac{1}{\beta}\left(1 \vee (\frac{1}{2\beta})^{\frac{1}{\beta}-1}\right), \tag{B.2}$$

for all $\theta\in(0,1]$.

C Integral estimates

The following is a trivial result that will come in useful.

Lemma C.1. *For finite $\delta > 0$, and $s < r \leq R < \infty$, $\int_s^r (R-t)^{\delta-1}dt \leq \frac{1}{\delta}(r-s)(T-s)^{\delta-1}$.*

Proof. Direct computation of the integral term yields

$$\int_s^r \frac{dt}{(R-t)^{1-\delta}} = \frac{1}{\delta}\{(R-s)^\delta - (R-r)^\delta\} \leq \frac{1}{\delta}\Big\{\frac{R-s}{(R-s)^{1-\delta}} - \frac{R-r}{(R-s)^{1-\delta}}\Big\} = \frac{r-s}{\delta(R-s)^{1-\delta}}.$$

$\square$

The following three lemmas and their proofs can be found in Section 2.1 of [GT13a]; the results on the integrals are proved exactly as the results on the sums.

Lemma C.2. *Let $\delta, \rho \in (0,1]$. Then for $B_{\delta,\rho} := \int_0^1 (1-r)^{\delta-1} r^{\rho-1} dr$, for any $0 \leq s < t \leq T$,*

$$\int_t^s \frac{dr}{(s-r)^{1-\delta}(r-s)^{1-\rho}} \leq B_{\delta,\rho}(s-t)^{\delta+\rho-1}.$$

Moreover, on the time-grid $\pi^{(\beta)} = \{0 = t_0 < \ldots < t_N = T \ : \ t_i = T - T(1-i/N)^{1/\beta}\}$, for any $0 \leq i < k \leq N$,

$$\sum_{j=i+1}^{k-1} (t_k - t_j)^{\delta-1}(t_j - t_i)^{\rho-1}\Delta_j \leq 2B_{\delta,\rho}(t_k - t_i)^{\delta+\rho-1}.$$

Lemma C.3. *Let $\delta \in (0, 1/2]$, $\rho > 0$ and $t \in [0,T)$. Suppose that, for a positive constant C_u, the finite positive real functions $u : [t,T] \mapsto [0,\infty)$ and $w : [t,T] \mapsto [0,\infty)$ satisfy*

$$u_t \leq w_t + C_u \int_t^T \frac{u_r dr}{(T-r)^{\frac{1}{2}-\delta}(r-t)^{\frac{1}{2}-\rho}}. \tag{C.1}$$

Then, for constants $\mathcal{C}_{(C.2a)}$ and $\mathcal{C}_{(C.2b)}$ depending only on C_u, T, δ and ρ,

$$u_t \leq \mathcal{C}_{(C.2a)} w_t + \mathcal{C}_{(C.2a)} \int_t^T \frac{w_r dr}{(T-r)^{\frac{1}{2}-\delta}(r-t)^{\frac{1}{2}-\rho}} + \mathcal{C}_{(C.2b)} \int_t^T \frac{u_r dr}{(T-r)^{\frac{1}{2}-\delta}}. \tag{C.2}$$

Moreover, on the time-grid $\pi^{(\beta)} = \{0 = t_0 < \ldots < t_N = T \ : \ t_i = T - T(1-i/N)^{1/\beta}\}$, suppose that the real functions $U : \pi^{(\beta)} \mapsto [0,\infty)$ and $W : \pi^{(\beta)} \mapsto [0,\infty)$ satify

$$U_i \leq W_i + C_u \sum_{j=i+1}^{N-1} \frac{U_j \Delta_j}{(T-t_j)^{\frac{1}{2}-\delta}(t_j - t_i)^{\frac{1}{2}-\rho}} \tag{C.3}$$

for all $i \in \{0, \ldots, N-1\}$. It follows that

$$U_i \leq 2\mathcal{C}_{(C.2a)} W_i + 2\mathcal{C}_{(C.2a)} \sum_{j=i+1}^{N-1} \frac{W_j \Delta_j}{(T-t_j)^{\frac{1}{2}-\delta}(t_j - t_i)^{\frac{1}{2}-\rho}} + 2\mathcal{C}_{(C.2b)} \sum_{j=i+1}^{N-1} \frac{U_j \Delta_j}{(T-t_j)^{\frac{1}{2}-\delta}}$$

for all $i \in \{0, \ldots, N-1\}$.

Lemma C.4. *Let $\delta \in (0, 1/2]$, $\rho > 0$ and $t \in [0,T)$. Suppose that the finite positive real functions $u : [t,T] \mapsto [0,\infty)$ and $w : [t,T] \mapsto [0,\infty)$ satisfy* (C.2) *for some positive constants $\mathcal{C}_{(C.2a)}$ and $\mathcal{C}_{(C.2b)}$. Then, for $\nu > 0$, there is a positive constant $\mathcal{C}^{(\nu)}$ (depending only on $\mathcal{C}_{(C.2a)}, \mathcal{C}_{(C.2b)}, T, \delta, \rho, \nu$) such that*

$$\int_t^T \frac{u_r dr}{(T-r)^{\frac{1}{2}-\delta}(r-t)^{1-\nu}} \leq \mathcal{C}^{(\nu)} \int_t^T \frac{w_r dr}{(T-r)^{\frac{1}{2}-\delta}(r-t)^{1-\nu}} \tag{C.4}$$

Moreover, on the time-grid $\pi^{(\beta)} = \{0 = t_0 < \ldots < t_N = T \;:\; t_i = T - T(1-i/N)^{1/\beta}\}$, suppose that the real functions $U : \pi^{(\beta)} \mapsto [0,\infty)$ and $W : \pi^{(\beta)} \mapsto [0,\infty)$ satisfy (C.3) *for all $i \in \{0,\ldots,N-1\}$. It follows that*

$$\sum_{j=i+}^{N-1} \frac{U_j \Delta_j}{(T-t_j)^{\frac{1}{2}-\delta}(t_j-t_i)^{1-\nu}} \leq 2\mathcal{C}^{(\nu)} \sum_{j=i+1}^{N-1} \frac{W_j \Delta_j}{(T-t_j)^{\frac{1}{2}-\delta}(t_j-t_i)^{1-\nu}}$$

for all $i \in \{0,\ldots,N-1\}$.

D Regularity results for inverse matrices

Lemma D.1. *Let $\xi > 0$ be finite and $A : \mathbb{R}^n \to \mathbb{R}^{l\times l}$ be symmetric and such that $\eta^\top A(x)\eta \geq \xi|\eta|^2$ for all $x \in \mathbb{R}^n$ and $\eta \in \mathbb{R}^l$. Then, for every $x \in \mathbb{R}^d$, the matrix $A(x)$ is invertible and $|A^{-1}(x)| \leq 1/\xi$. Moreover, if $x \mapsto A(x)$ is γ-Hölder continuous,then it's inverse $x \mapsto A^{-1}(x)$ is also γ-Hölder continuous.*

Proof. Due to the condition $\eta^\top A(x)\eta \geq \xi|\eta|^2$, it follows that $A(x)$ is positive definite for every $x \in \mathbb{R}^n$. This implies that the singular values of $A(x)$ are all greater than ξ [GVL96, Theorem 8.1.2], and so $A(x)$ is invertible. Using the singular value decomposition of $A(x)$ to construct the inverse as in [GVL96, Section 5.5.4], the maximal sigular value of $A^{-1}(x)$ is less than $1/\xi$ and so, using [GVL96, Section 2.5.2] combined with the singular value decomposition of $A^{-1}(x)$, the matrix 2-norm of $A^{-1}(x)$ is equal to its maximal singular value, i.e. $|A^{-1}(x)| \leq 1/\xi$ for all $x \in \mathbb{R}^d$. Now, let x and y be elements in $\mathbb{R}^d$. Since $A^{-1}(y) - A^{-1}(x)$ is equal to

$$-A(x)^{-1}\big(A(y) - A(x)\big)A(y)^{-1},$$

it follows that

$$|A^{-1}(y) - A^{-1}(x)| \leq |A^{-1}(x)||A(y) - A(x)||A^{-1}(y)| \leq \frac{L_A}{\xi^2}|x-y|^\gamma,$$

where L_A is the Hölder constant of A. □

Proof of Lemma 4.1. Let $t \in [0,T)$ be fixed, and define $A : [0,T) \times \mathbb{R}^d \to \mathbb{R}^{d\times d}$ by $A(t,x) = \sigma(t,x)\sigma(t,x)^\top$. It can be computed directly that $\sigma^{-1}(\cdot) = \sigma(\cdot)^\top A^{-1}(\cdot)$, whether or not d equals q. It follows from uniform ellipticity ($\mathbf{A}_{\text{u.e.}}$) and Lemma D.1 that $|A^{-1}(t,x)| \leq 1/\bar{\beta}$ for all $(t,x) \in [0,T) \times \mathbb{R}^d$. Due to the differentiability condition ($\mathbf{A}_{\mathbf{b},\sigma}$) on $\sigma(t,\cdot)$, $\sigma(t,\cdot)$ is Lipschitz continuous uniformly in t with Lipschitz constant $\|\nabla_x\sigma\|_\infty$, and, using additionally the equality $A(t,x) - A(t,y) = \sigma(y)(\sigma(t,x)^\top - \sigma(t,y)^\top) + (\sigma(t,x) - \sigma(t,y))\sigma(t,x)^\top$, $A(t,\cdot)$ is Lipschitz continuous uniformly in t with Lipschitz constant $2\|\sigma\|_\infty\|\nabla_x\sigma\|_\infty$. Using Lemma D.1, it follows that $A^{-1}(t,\cdot)$ is Lipschitz continuous uniformly in t with Lipschitz constant $2\|\sigma\|_\infty\|\nabla_x\sigma\|_\infty/\bar{\beta}^2$. For any $(x,y) \in (\mathbb{R}^d)^2$, $\sigma(t,x)^{-1} - \sigma(t,y)^{-1}$ is equal to $\big(\sigma(t,x)^\top - \sigma(t,y)^\top\big)A^{-1}(t,x) + \sigma(t,y)^\top\big(A^{-1}(t,x) - A^{-1}(t,y)\big)$. and therefore

$$|\sigma(t,x)^{-1} - \sigma(t,y)^{-1}| \leq \frac{\|\nabla_x\sigma\|_\infty}{\bar{\beta}}|x-y| + \frac{2\|\sigma\|_\infty\|\nabla_x\sigma\|_\infty}{\bar{\beta}^2}|x-y|.$$

The proof that $\sigma^{-1}(\cdot,x)$ is 1/2-Hölder continuous is essentially the same and we do not include it. □

References

[BDH+03] P. Briand, B. Delyon, Y. Hu, E. Pardoux, and L. Stoica. $\mathbf{L}_p$ solutions of backward stochastic differential equations. *Stochastic Processes and their Applications*, 108(1):109–129, 2003.

[BL13] P. Briand and C. Labart. Simulation of BSDEs by Wiener Chaos Expansion. *To appear in Annals of Applied Probability*, 2013.

[BT04] B. Bouchard and N. Touzi. Discrete time approximation and Monte Carlo simulation of backward stochastic differential equations. *Stochastic Processes and their Applications*, 111:175–206, 2004.

[CD12] D. Crisan and F. Delarue. Sharp derivative bounds for solutions of degenerate semi-linear partial differential equations. *Journal of Functional Analysis*, 263(10):3024–3101, 2012.

[CR14] J. F. Chassagneux and A. Richou. Numerical simulation of quadratic BSDEs. *Available on http://arxiv.org/abs/1307.5741*, 2014.

[DG06] F. Delarue and G. Guatteri. Weak existence and uniqueness for forward-backward SDEs. *Stochastic Processes and their Applications*, 116(12):1712–1742, 2006.

[EKPQ97] N. El Karoui, S. Peng, and M. C. Quenez. Backward stochastic differential equations in finance. *Mathematical Finance. An International Journal of Mathematics, Statistics and Financial Economics*, 7(1):1–71, 1997.

[FJ12] S. J. Fan and L. Jiang. $\mathbf{L}_p$ solutions of finite and infinite time interval BSDEs with non-Lipschitz coefficients. *Stochastics*, 84(4):487–506, 2012.

[Fri64] A. Friedman. *Partial differential equations of parabolic type.* Prentice-Hall Inc., Englewood Cliffs, N.J., 1964.

[GGG12] C. Geiss, S. Geiss, and E. Gobet. Generalized fractional smoothness and $\mathbf{L}_p$-variation of BSDEs with non-Lipschitz terminal condition. *Stochastic Processes and their Applications*, 122(5):2078–2116, 2012.

[GL06] E. Gobet and J. P. Lemor. Numerical simulation of BSDEs using empirical regression methods: theory and practice. In *Proceedings of the Fifth Colloquium on BSDEs (29th May - 1st June 2005, Shangai) - Available on http://hal.archives-ouvertes.fr/hal-00291199/fr/*, 2006.

[GL07] E. Gobet and C. Labart. Error expansion for the discretization of backward stochastic differential equations. *Stochastic Processes and their Applications*, 117(7):803–829, 2007.

[GL10] E Gobet and C. Labart. Solving BSDE with adaptive control variate. *SIAM Journal on Numerical Analysis*, 48(1):257–277, 2010.

[GM+05] E. Gobet, R. Munos, et al. Sensitivity analysis using Itô-malliavin calculus and martingales, and application to stochastic optimal control. *SIAM Journal on control and optimization*, 43(5):1676–1713, 2005.

[GM10] E. Gobet and A. Makhlouf. $\mathbf{L}_2$-time regularity of BSDEs with irregular terminal functions. *Stochastic Processes and their Applications*, 120(7):1105–1132, 2010.

[GT13a] E. Gobet and P. Turkedjiev. Approximation of backward stochastic differential equations using malliavin weights and least-squares regression. *To appear in Bernoulli, available on http://hal.archives-ouvertes.fr/hal-00855760*, 2013.

[GT13b] E. Gobet and P. Turkedjiev. Linear regression MDP scheme for discrete backward stochastic differential equations under general conditions. *Available on http://hal.archives-ouvertes.fr/hal-00642685*, 2013.

[GVL96] G. H. Golub and C. F. Van Loan. *Matrix computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, third edition, 1996.

[HIM05] Y. Hu, P. Imkeller, and M. Müller. Utility maximization in incomplete markets. *The Annals of Applied Probability*, 15(3):1691–1712, 2005.

[HNS11] Y. Hu, D. Nualart, and X. Song. Malliavin calculus for backward stochastic differential equations and application to numerical solutions. *Ann. Appl. Probab.*, 21(6):2379–2423, 2011.

[IDR10] P. Imkeller and G. Dos Reis. Path regularity and explicit convergence rate for BSDE with truncated quadratic growth. *Stochastic Processes and their Applications*, 120(3):348–379, 2010.

[JS03] J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 2003.

[Kus03] S. Kusuoka. Malliavin calculus revisited. *The University of Tokyo. Journal of Mathematical Sciences*, 10(2):261–277, 2003.

[MZ02] J. Ma and J. Zhang. Representation theorems for backward stochastic differential equations. *The Annals of Applied Probability*, 12(4):1390–1418, 2002.

[Nee11] C. Nee. *Sharp Gradient Bounds for the Diffusion Semigroup*. PhD thesis, Imperial College London, 2011.

[Nua06] D. Nualart. *The Malliavin calculus and related topics*. Probability and its Applications (New York). Springer-Verlag, Berlin, second edition, 2006.

[REK00] R. Rouge and N. El Karoui. Pricing via utility maximization and entropy. *Mathematical Finance. An International Journal of Mathematics, Statistics and Financial Economics*, 10(2):259–276, 2000. INFORMS Applied Probability Conference (Ulm, 1999).

[Ric11] A. Richou. Numerical simulation of BSDEs with drivers of quadratic growth. *The Annals of Applied Probability*, 21(5):1933–1964, 2011.

[Ric12] A. Richou. Markovian quadratic and superquadratic BSDEs with an unbounded terminal condition. *Stochastic Processes and their Applications*, 122(9):3173–3208, 2012.

[RY99] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.

[Zha04] J. Zhang. A numerical scheme for BSDEs. *The Annals of Applied Probability*, 14(1):459–488, 2004.

[Zha05] J. Zhang. Representation of solutions to BSDEs associated with a degenerate FSDE. *The Annals of Applied Probability*, 15(3):1798–1831, 2005.